Biotechnology Division - Table of Contents

Division Overview

The Biotechnology Division is the focus of the NIST effort addressing critical measurement and data needs for other government agencies, and the rapidly developing biotechnology industry.

MISSION

The mission of the Division is to provide the measurement infrastructure necessary to advance the commercialization and application of biotechnology. This is achieved by developing a scientific and engineering technical base along with reliable measurement techniques and data to enable U.S. industry to produce biochemical products, and to enable the government to apply advances in biotechnology to the benefit of societal needs. The Division has established a variety of long-range research projects to maintain critical expertise needed for the development of Standard Reference Materials, Standard Reference Databases, and advanced measurement methods. The Division fosters collaboration among NIST scientists conducting biology-related research, and strives to reach the goals set by the NIST Chemical Science and Technology Laboratory strategic plan.

Division scientists participate in scientific meetings, topical workshops, and numerous national and international organizations such as: Biotechnology Industry Organization (BIO), IUPAC Commission on Biophysical Chemistry, ASTM Committee E-48 on Biotechnology, the International Measurement Standards Consultative Committee for the Amount of Substance (CCQM), Bioanalytical Working Group. Division members were also active as reviewers for the NIST Advanced Technology Program (ATP), for several NSF and DOE programs, NIH study section panels, for the Department of Homeland Security on issues related to bioterrorism defense, and with the Department of Justice on issues related to forensics and human identification.

The staff of the Biotechnology Division consists of 36 NIST employees and a comparable number of contract researchers, guest scientists, and post-doctoral fellows. The Division is organized into four groups: (1) **DNA Technologies**; (2) **Bioprocess Measurements**; (3) **Structural Biology**; and (4) **Bimolecular Materials**. Brief descriptions of technical highlights from each Group are given below.

Selected Program Highlights:

DNA TECHNOLOGIES

The DNA Technologies Group has research projects that are included in the Program Areas of Health and Medical Products, Forensics and Homeland Security, and Food and Nutritional Products. **Advanced mass spectrometry measurements of DNA damage** are used to describe the cellular accumulation of two major oxidative stress-induced DNA lesions in cells of Cockayne syndrome (CS) patients after exposure to ionizing radiation. As a disease with implications for understanding the human aging process, these studies are undertaken as a collaborative effort with scientists at the National Institute of Aging. Projects in the area of **DNA diagnostics for the detection of human disease** include the NIST-National Cancer Institute Biomarkers Validation Laboratory (BVL), the NIST component of the Early Detection Research Network which serves to refine recently discovered cancer biomarkers, and to format new

research tests for field trials in EDRN clinical laboratories. Another area is the study of cellular biomarkers that can be used for quality assurance of tissue-engineered medical products in terms of genetic damage. In the **human identity/forensic science** project, the group focuses on new methods for DNA profiling, ranging from developing well-characterized DNA standards for restriction fragment length polymorph-isms (RFLPs) to performing research for rapid determination of DNA profiles by polymerase chain reaction (PCR) amplification and automated detection of fragments. New methods were developed for identification of victims of the World Trade Center (WTC) disaster of September 11, 2001 where the high degree of DNA fragmentation due to the severe environmental conditions has meant that only about 50% of the specimens yielded results with standard DNA testing methods.

BIOPROCESS MEASUREMENTS

The Bioprocess Measurements Group (**http://www.cstl.nist.gov/div831/bioprocess/**) is concerned with the development of measurement methods, databases, and generic technologies related to the use of biomolecules and biomaterials. The results are directed at the biomanufacturing, agbiotechnology, and pharmaceutical industries and, most recently, to Homeland Defense, where there are needs for the detection and quantification of very small amounts of biological materials. The effort is organized into four project areas that are part of the CSTL Pharmaceuticals and Biomanufacturing, Food and Nutritional Products, and Forensics and Homeland Security Programs. In the **spectroscopy of biological systems** project, one study is directed at investigating the mechanism of fluorescence resonance energy transfer (FRET) when it is used to quantify the extent of a polymerase chain reaction (PCR). In the figure, FRET efficiency is seen to decline by five-fold as a function of fluorophore separation, counted as number of nucleic bases between the fluorophore labeling sites. These results are relevant to an ongoing project done in collaboration with the USDA to detect and measure genetic modifications in grain, sometimes referred to as GMOs. In the **biocatalysis** project, enzyme characterization is being carried out to address industrially important biotransformation problems such as those found in hydroxylation and aromatic amino acid metabolic pathways. The methods used include site-directed mutagenesis, circular dichroism, ellipsometry, spectroelectro-chemistry, and X-ray diffraction to characterize several key steps along metabolic pathways. In the **biothermodynamics** project, chromatography and microcalorimetry are used with chemical equilibrium analysis of complex reacting systems to develop thermodynamic data for industrially important biotransformations that are included in the NIST Standard Reference Database "Thermodynamics of Enzyme-catalyzed Reactions." A new project, **microbial forensics**, has recently started to develop standard methods, materials and data related to the national efforts to defend against threats of biological warfare.

STRUCTURAL BIOLOGY

The Structural Biology Group at the Center for Advanced Research in Biotechnology (CARB) is focused in key areas of industrial biotechnology, especially in the Pharmaceuticals and Biomanufacturing Program. These areas are supported at CARB through a highly interactive group of scientists, from both the University of Maryland Biotechnology Institute (UMBI) and NIST. In the **structure-function of biological macromolecules**, a recent study of the dimerization of two homologous strands of genomic RNA reveals an essential reaction in the replication of retroviruses such as HIV-1 (see figure). A new effort has been launched in structural genomics that capitalizes on existing expertise macromolecular structure determination

by X-ray crystallography. This project area also supports the NIST activity related to the Protein Data Bank (**http://pdb.nist.gov/**). Results from the physical, molecular and cellular biochemistry studies of key recognition elements in G coupled protein receptors suggest new, quantitative models for signal transduction pathways in vision and viral infection. The energetics of enzyme-catalyzed reactions are being studied by differential stopped flow microcalorimetry. A recent highlight was completion of a study of the binding interactions in the model system composed of a well-characterized enzyme-inhibitor pair, namely bovine carbonic anhydrase II (CA II) and 4-carboxybenzenesulfonamide (CBS). Under the initiative of The Molecular Interactions Research Group (MIRG) of the Association of Biomolecular Resource Facilities (ABRF), this model system for monitoring complex formation was distributed to a panel of analytical ultracentrifugation (AUC), isothermal titration calorimetry (ITC), and surface plasmon resonance (SPR). operators. The study participants were asked to measure one or more of the following: 1) the molecular mass, homogeneity, and assembly state of CAII by AUC; 2) the affinity and thermodynamics for complex formation by ITC; and 3) the affinity and kinetics of complex formation by SPR. The results of the study showed excellent agreement between the ITC and SPR results on the binding thermodynamics for complex formation. The AUC results showed that the enzyme exists as a monomer in solution in agreement with the ITC results. The results from this study provide a benchmark for comparing the capabilities of individual laboratories and defining the utility of the different instrumentation.

From the **bioinformatics for disease markers** project, an HIV structural database has been developed that was highlighted in Science magazine's Netwatch (Science, May 30, 2003; srdata.nist/hivdb). A project on **gene expression** is just underway in collaboration with platform developers, reagent makers, clinical users and other government agencies (Nature, January 8, 2004, p. 91; http://www.cstl.nist.gov/biotech/workshops/ERCC2003/).

BIOMOLECULAR MATERIALS

The Biomolecular Materials Group studies the behavior of biological molecules and adapts them for novel technological and scientific applications, and to emerging needs of bioterrorism research. Measurement methods including surface plasmon resonance, IR spectroscopy, ellipsometry, electrophysiology, impedance spectroscopy, chemical synthesis, atomic force microscopy, and confocal microscopy are combined with computer simulations to carry out projects in nanobiotechnology, tissue engineering, and mitochondrial proteomics. In the nanobiotechnology project, single nanometer-scale pores were employed to study biological transport processes, to read information within single biomolecules and to detect multiple analytes in solution. In the tissue engineering project, the need for biomarkers, physical standards and measurement technologies for tissue engineering are being addressed to assure quality control during manufacturing and storage of engineered medical products. A method for reproducibly and reliably fabricating films of collagen to provide surfaces on which cells can be grown is under development. Data shown in the figure illustrate that the films can induce morphological changes in vascular smooth muscle cells that mimic their response in healthy and diseased arteries. The ability to characterize these films with surface analytical techniques permits the evaluation of how changes in the collagen substrate influence cellular responses, potentially leading to reference materials. In another study, the ability of chemokines to interact with G-proteins, which are important target molecules of the pharmaceutical industry, is interrogated with surface plasmon resonance spectroscopy. Cell fragments containing the transmembrane proteins are immobilized to a solid support where cells containing the G-protein

CCR5 bind to a surface. An antibody against CCR5 subsequently binds to these membranes, but not to a control surface. These results suggest that the technique may be useful for detecting the binding of small ligands to these receptors. The emphasis of the mitochondrial proteomics project will be to address needs of the mitochondrial and proteomics communities as outlined in a September 2002 workshop and to develop general protocols for handling and characterizing membrane associated proteins.

The Short Tandem Repeat DNA Internet Database:
http://www.cstl.nist.gov/biotech/strbase/

Protein Data Bank:
http://rcsb.nist.gov/

Thermodynamics of Enzyme-Catalyzed Reactions:
http://wwwbmcd.nist.gov:8080/enzyme/enzyme.html

The Biological Macromolecule Crystallization Database:
http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html

HIV Protease Database:
http://srdata.nist.gov/hivdb/

A Metric of Amino Acid Exchangeability Based on Experimental Data

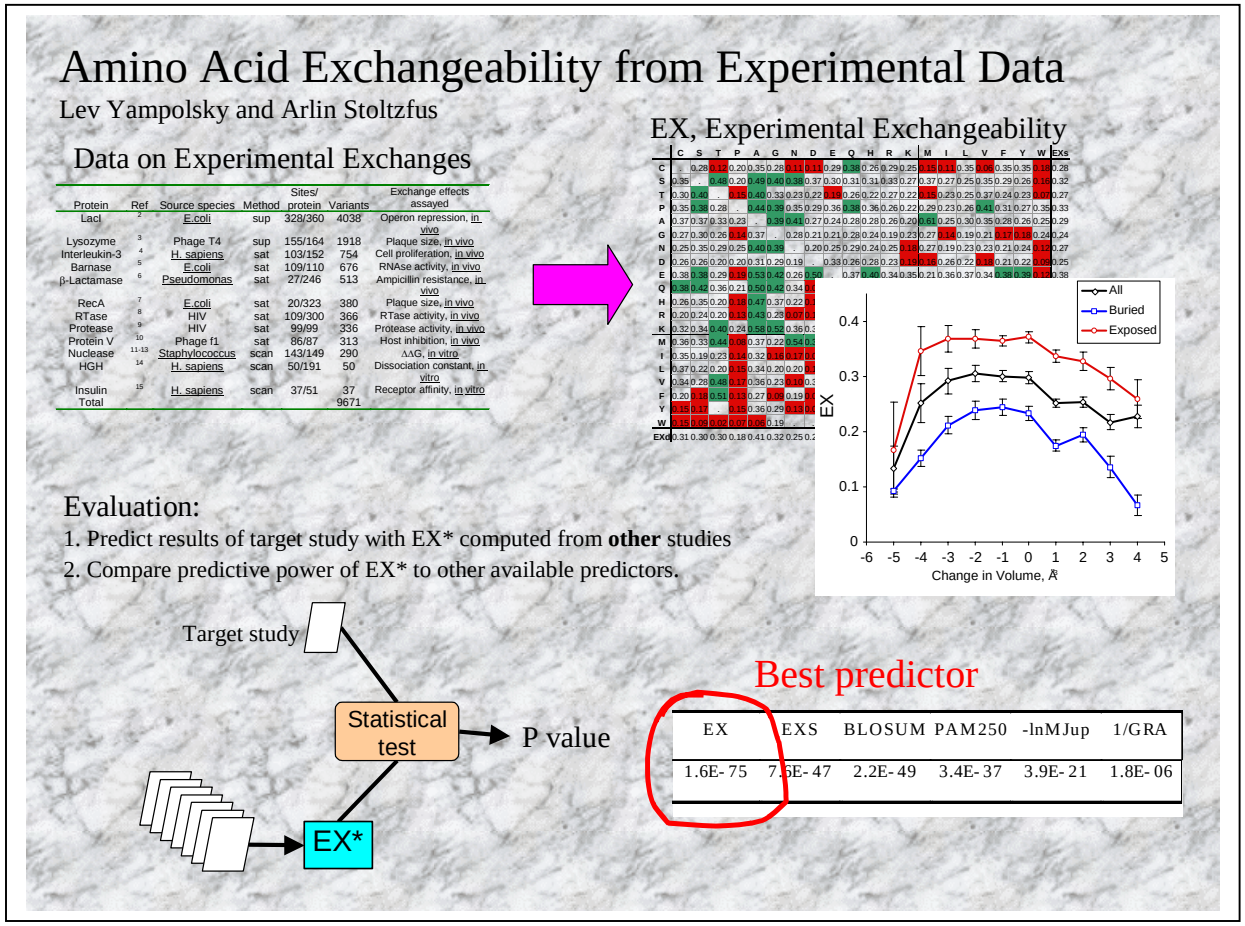
Authors: ***Arlin Stoltzfus (831), Lev Y. Yampolsky (University of Maryland Biotechnology Institute)***

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: Comparative sequence analysis, the analysis of the differences between similar but non-identical genes or proteins, is a fundamental part of computational genomics, necessary for genome annotation and functional prediction in bioinformatics. The significance of amino acid sequence differences between proteins is often interpreted by means of a classification of amino acids (e.g., polar vs. non-polar), or by some continuous measure of amino acid distance. Available measures based on observed propensities of evolutionary change are undesirable because they include confounding mutational effects; physicochemical measures are uncertain because they rely on some simplified model of the role of amino acids in proteins. What is needed is a measure that reliably reflects this role, without the confounding effects of mutation. Such a measure has been derived from published results of 9671 experimental amino acid replacements in 12 proteins. Statistical cross-validation demonstrates the success of this method. The resulting measure of amino acid similarity improves upon available alternatives, whether used as a measure of the severity of laboratory mutants, the severity of naturally occurring mutations in humans, or the acceptance probability of evolutionary changes.

Purpose: Comparative analysis of protein sequences relies on interpreting differences that arise primarily by the evolutionary replacement of one amino acid by another. How will such changes alter protein activity? Probabilities of evolutionary change reflect, not only effects on the protein, but also rates of mutation between codons. For example, the low rate of evolutionary interchange of Methionine and Cysteine might be due to incompatibility at the protein level, but is more likely explained by the genetic code, which dictates that three separate mutations are needed to go from a Methionine codon to a Cysteine codon. Effective comparative sequence analysis relies on the ability to separate these two factors: the exchangeability of amino acids and the mutational exchangeability of codons.

Accomplishments: Prior work on this project resulted in the identification of systematic data on the effects of switching amino acids in proteins; the development of a statistical method for transforming the results from different studies to a common scale; the derivation of a measure called the "experimental exchangeability", the mean effect $\mathbf{EX_{ij}}$ of replacing amino acid **i** with amino acid **j**, based on data from 9671 experimental amino acid exchanges from 12 selected studies; and a test for evaluating the power of an exchangeability metric, which was used to show that the method for computing EX works. The first phase of the project was completed this year, marked by the submission of a manuscript for publication and the release of the EX matrix to researchers in the field of protein sequence analysis. Also this year, the second phase of the project has begun, with two demonstration projects to assess the utility of EX in protein sequence analysis. *The first project addresses the disease-causing effects of amino acid changes in human genes.* For each type of amino acid change, the observed disease-causing propensity is computed from data on known disease mutants and SNPs (single-nucleotide polymorphisms). EX outperforms other possible metrics of severity, accounting for 46 % of the variance in the logarithm of disease-causing potential (see figure).



Impact: The impact of this work is hard to judge at present, because EX is only now being released, and because most current methods of sequence analysis are not designed to take advantage of the distinctive properties of EX (see Future Plans).

Future plans: Future plans involve more extensive modeling and analysis both of protein mutants and of evolution, and the modification of methods of sequence analysis to take advantage of an asymmetric measure of pure exchangeability with statistical weights. The evolutionary analysis will determine the relative contributions of mutational and selective factors in accounting for patterns in the evolution of proteins and of protein-coding genes. The analysis of mutants will apply simple models for predicting frequencies of disease-causing alleles, and will address the long-term independence of mutational and selective factors, as well as relate activity effects to thermodynamic effects on stability of mutants.

Mapping the Protective Antigen Binding Site on the Anthrax Toxin Receptor

Author: **Kevin D Ridge (831)**

CSTL Program: Forensics and Homeland Security, FY03 Exploratory Research Project

Abstract: Anthrax is a disease caused by the gram-positive, spore forming bacterium *Bacillus anthracis*. Humans are typically incidental hosts through ingestion of contaminated foods of animal origin, various animal products, and more recently, malicious exposure. Because of the heightened public awareness and concern surrounding anthrax, there is an acute need to understand the molecular mechanisms governing the intoxication process and to apply this knowledge to the development of new detection methods and therapeutic strategies that confront this disease. The major contributing factor to the pathogenesis of *Bacillus anthracis* is a secreted, three-component toxin. The three proteins comprising the toxin are the lethal factor (LF), the edema factor (EF), and the protective antigen (PA). Although the mechanism of anthrax intoxication is not entirely clear, some discrete steps have been delineated through a variety of experimental approaches. Initially, PA binds to a cell surface integral membrane protein, the anthrax toxin receptor (ATR), and is cleaved by a cell surface protease generating a truncated carboxyl-terminal form of the protein (PA63). PA63 possesses the inherent ability to self assemble and the oligomeric form of the protein is able to bind EF or LF with high affinity. The resulting complexes are then internalized where they exert their toxic effects.

Purpose: A comprehensive understanding of the molecular mechanisms governing the pathogenesis of *Bacillus anthracis* requires quantitative information about the interaction of PA with its host cell membrane bound receptor, ATR. As the binding of PA to ATR can be considered as the first step in the intoxication process, it represents a potentially important target site for early detection and therapeutic intervention. However, since little is currently known about this interaction at the molecular level, mapping of the PA binding site on ATR is of intense interest to the scientific community as well as those charged with protecting public health.

Major Accomplishments: The binding site for PA on ATR is being identified through a combination of domain and mutagenesis approaches. As a first step, the gene for ATR was obtained by PCR of CHO-K1 cell genomic DNA. Initially, several different PCR products of various lengths (0.5 – 1.4 kb) were obtained using different sets of amplifying primers. This approach ultimately yielded a full-length ATR gene (1.2 kb) that shares high DNA sequence homology with another related integral membrane protein, TEM8. However, due to the numerous redundant restriction endonuclease sites in the native ATR gene that would hamper mutational analysis, an alternative approach also was undertaken to construct a synthetic ATR gene. For this purpose, a strategy for the design of an ATR gene with evenly spaced, unique restriction endonuclease cleavage sites was pursued and corresponding DNA fragments synthesized that allow for assembly and expression of a complete synthetic ATR gene.

Impact: Mapping the PA binding site(s) on ATR is necessary for advancing our understanding of these biological macromolecules and their mechanistic role in the pathogenesis of *Bacillus anthracis*. Clearly, knowledge of the critical “establishment” stages of anthrax intoxication should provide decisive targets for the early detection of anthrax and subsequent intervention protocols responding to toxin exposure.

Future Plans**:** The availability of the ATR gene enables a systematic study of the PA binding site. A synthetic PA gene has been constructed by Edward Eisenstein of UMBI/CARB who is investigating PA63 assembly. With these proteins in hand, detailed biochemical and biophysical measurements of the ATR/PA interaction are likely to provide a quantitative description of the anthrax intoxication process.

Standards for Detection of Biological Threat Agents

Authors: *Kenneth D. Cole, L. Wang, A. Gaigalas, and D. Hancock (831)*

CSTL Program: Forensics and Homeland Security

Abstract: We are working with the Department of Homeland Security (DHS) and other government agencies to determine the best ways that NIST can meet the needs for data, measurements, and reference materials for the detection of biological threat agents. We are now in the process of charactering surrogates that can safely be used for the testing of detection devices, remediation technologies, and personnel training. There are many instances where these surrogates are preferable to the use of the actual bacteria, viruses, and toxins. We have identified several surrogates and are in the process of developing methods to fully characterize them. We will use these well-characterized surrogates as model systems to study and improve the instruments and devices used for the detection of biological threats. If necessary, follow-up work will be performed at BSL-3 laboratories that are able to handle the live agents. Additionally, the detection of biological threats is done from environmental samples that represent serious challenges because of their complex and variable compositions and the possibility of interferences.

Purpose: The purpose of the project is to identify the needs and customers for standards needed for the detection of biological threat agents. We are working in close consultation with other government labs and agencies to identify these needs and opportunities for NIST. This is being done through the task forces (sponsored by the DHS) and by collaborations we have established and are still establishing. We are identifying the best ways that NIST can assist the DHS to provide measurements, data, and materials to provide national security.

Major Accomplishments: We have identified several surrogates for the biological threat agents that we can safely use in our laboratories. The surrogates include a spore-forming bacteria (Bacillus globigii), an RNA virus (the bacteriophage MS2), and the Ricin A chain. Detection of biological threats currently depends upon either recognition by antibodies or detection of specific nucleic acid sequences. We are in the process of developing improved sensitivity and reliability of the assays. The methods we are studying include classical microbiological techniques, flow cytometry of fluorescent-labeled surrogates, fluorescence microscopy, and DNA based assays such as PCR.

Impact: The development of methodologies to characterize surrogates for biological threats is important and challenging because of the needs of accuracy, sensitivity, and reliability. It is essentially certain that well-characterized surrogates will be used for standardizing the devices and instruments used for the detection and measurement of these threats.

Future Plans: We will continue to build our strengths in characterizing these surrogates and identifying new ones. The devices and instruments need to be tested using realistic surrogates and from complex environmental samples. NIST can help assure national security by helping to test these devices and proving data and measurements. The surrogates will be used to study the devices and instruments used to detect existing as well as new emerging threats. These complex organisms and biological molecules will require sophisticated techniques to characterize them. New devices and instruments that are being developed will also need well-characterized surrogates to determine their performance.

Biological Macromolecule Crystallization Database

Authors: *Gary Gilliland, Michael Tung, TN Bhat, and Jane Ladner(831)*

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: The NIST/CARB Biological Macromolecule Crystallization Data (BMCD), NIST Standard Reference Database21, contains crystal data and crystallization conditions for biological macromolecules. The BMCD was designed as a tool to assist x-ray crystallographers in the development of protocols to crystallize biological macromolecules, those that have previously been crystallized, and those that have not. Described here is recent work that has been carried out that was required to improve the ability to update the resource and a discussion of what future developments will be implemented for the BMCD.

Purpose: The Biological Macromolecule Crystallization Database (BMCD) includes the crystallization data for peptides, proteins, protein-protein complexes, nucleic acids, nucleic acid-nucleic acid complexes, protein-nucleic acid complexes, and viruses for which diffraction quality crystals have been obtained and reported in the literature. Besides collecting data from the literature, data can be acquired from other resources such as the Protein Data Bank. Therefore, a mapping between BMCD data model and the mmCIF (macromolecular Crystallographic Information File) data representation of PDB entries was carried out.

Major Accomplishments: The mapping between BMCD data model and mmCIF data representation of PDB entries resulted in a new and expanded data representation for the BMCD. This new data representation gives a consistent way to carry out data exchange between the BMCD and other structural biology resources, and hence provides a new avenue for BMCD data acquisition. This new data representation is not only compatible with the earlier BMCD data conventions, but also with the PDB data. The data model simplifies and reduced the number of many-to-many relationships making the database much easier for data management. For example, tables for depositors and auditing information have been added. With this new data model, data to be added to the BMCD can be extracted from the PDB mmCIF data set using a Star (CIF) Perl API.

Impact: Through the efforts described above and through many discussions with structural biologists, the new data model for the BMCD has resulted in improved data management capabilities. With this new data model, data to be added to the BMCD can be extracted from the PDB mmCIF data set using a Star (CIF) Perl API. The PDB resource contains more than 18,000 crystals for which data can be extracted. A significant fraction (15-20%) contains comprehensive crystallization information. Therefore, with this new data model data, the BMCD can be significantly expanded.

Future Plans: Even though we can obtain data from PDB entries, much of the crystallization data required to reproduce the experiments will still be lacking from the majority of entries. The missing information includes the crystallization method, crystal growth temperature and pH, and chemical additives. Rather than collecting this information from the literature, which is very time consuming, a data mining approach using data clustering will be carried out. Clustering will be

based on the characterization of crystals and molecules, using such data items as unit cell parameters and macromolecular or protein properties. A new BMCD website will be created that will deliver the results of clustering rather than just the individual entry information. Also the new site will allow more complex queries. The data management and relational model mentioned above is not ideal for these purposes because it emphasizes none-duplicate entries. A new generation data model called a “Data Mart” will be employed with the BMCD website to provide new ways of analyzing and querying the underlying data.

Desiccation-tolerance of Genomic DNA of the Cyanobacterium *Nostoc commune*

Authors: *H. Rodriguez, M. Birincioglu (831); P. Jaruga (UMBC, MD); B. Shirkey, S.C. Smith, D.J. Wright, M. Potts (Virginia Tech Center for Genomics and Virginia Tech, Blacksburg, VA)*

CSTL Program: Biomaterials, Project 831-002, Tissue Engineering

Abstract: Water is indispensable for the maintenance of cell integrity and function. The removal of water from a cell causes such severe perturbation that the cell may die within seconds of its exposure to the atmosphere. As a consequence, desiccation is likely to have been an important determinant in the evolution of primitive cells. Today, desiccation remains the most acute environmental stress suffered by living cells. The mechanisms that enable a cell to survive as it enters the air-dried state, as it remains desiccated for extended periods of time (decades or more), and as it leaves the air-dried state are unknown. The aim of this work is to gain a thorough understanding of the structural, physiological and molecular basis for desiccation-tolerance in this ecologically significant, cosmopolitan microorganism. Using the Cyanobacteria, *Nostoc commune*, as a model, we are studying the effects of several dry-down technologies to dry, stabilize and preserve cells. This work showed demonstrated the stability of desiccated DNA *in vivo*. The results of this study have been published in Nucleic Acids Research. Understanding these processes will aid the development of better long-term storage methods of tissue-engineered medical products.

Purpose: To identify the key physiological processes and metabolic networks that contributes to desiccation tolerance using functional genomics and proteomics.

Cryopreservation for Long-Term Storage of Tissue-Engineered Medical Products.

Major Accomplishments: This work showed demonstrated the stability of desiccated DNA *in vivo*. The results of this study have been published in Nucleic Acids Research.

Impact: This work is complimentary to the ongoing research project in the field of biomarkers of tissue engineering within the DNA Technologies Group. It broadens the work at NIST by directly applying the biomarkers developed for the field of tissue engineering to the field of cell desiccation tolerance and cryopreservation.

Future Plans: It is hoped that this study will be extended to other issues related to the cryopreservation of cells – thus, aiding the development of better long-term storage methods of tissue-engineered medical products.

Bioseparations - Separation of Different Topological Forms of Circular DNA

Author: K. D. Cole

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: The DNA from many microorganisms and cellular organelles is circular. Plasmids are important examples of circular DNA that are widely used to clone genes and express foreign proteins. We are studying the separation of circular DNA to improve methods for the analysis and the isolation of large amounts of plasmid DNA. Circular DNA can be selectively retained in gels by electric fields by a process termed electrophoretic trapping. We have studied the effects of the gel structure and the topological form of the DNA on trapping. The concentration of polymer (agarose) used to form the gel has a large effect on the formation of the "traps" for circular DNA. Lower concentration gels had a higher density of traps and the traps were more heterogeneous when compared to traps formed in gels formed with higher concentrations of agarose. Electrophoretic trapping can be used for very rapid separations and for the isolation of very pure circular DNA when thin layers are used as the media. An apparatus for thin-layer trapping was constructed and tested using circular plasmid DNA. The apparatus allows separations to be done in a few minutes and to thus obtain high purity circular DNA.

Purpose: Improved methods for the production, characterization, and purification of circular DNA are needed to meet the needs of the pharmaceutical, biotechnology, and molecular biology communities. We are studying electrophoretic trapping in gels as a process that separates DNA based on its physical form. Our efforts have been to characterize the process, develop models, and attempt to exploit electrophoretic trapping for improved analytical and preparative methods for circular DNA. The migration of circular DNA can be completely arrested in gels when the electric field strength reaches a critical value. The factors that determine the formation of the electrophoretic traps in gels are not well understood. A better understanding of these factors would allow us to control their formation and to design better gels for the selective separation of circular DNA.

Major Accomplishments:

1. The electric field strength for the onset of trapping, the apparent trap density, and the release characteristics of circles from traps were studied using both direct current and field inversion gel electrophoresis (FIGE) experiments. The average distances before 13 kbp (kilobase pairs) open circles were trapped were ~80 mm in 0.25% gels, 180 mm in 1% gels, and greater than 500 mm in 2.5% gels. The higher trap density and longer trap length in low concentration gels explains the experimental observations that under some conditions DNA circles migrate faster in higher concentration gels compared to their migration in lower concentration gels. Publication: Cole, K. D.; Åkerman, B., "The influence of agarose concentration in gels on the electrophoretic trapping of circular DNA," Separation Sci. Technol. 38, 2121-2136 (2003).

2. An apparatus was designed for the electrophoretic capture and recovery of circular DNA in thin layers (membranes). Rapid separations were done by the use of a low conductivity buffer and high electric field strengths. Two methods that specifically retain circular DNA in the membranes are demonstrated using the super coiled and open circular forms of two plasmids with sizes of 4.4 kbp and 13 kbp. Electrophoretic trapping (by an impalement mechanism) in

agarose gel-filled membranes used electric field strength to immobilize circular DNA in the membranes. These membranes and methods should be scaleable to process large number of samples or to prepare larger amounts of pure circular DNA by increasing the area of the membranes. The apparatus facilitates the rapid screening of a large number of trapping materials and methods for recovery of DNA. Publication: Cole, K. D. "An apparatus for electrophoretic trapping of circular DNA in thin layers," Biotechnol. Applied Biochem. 37, 251-257 (2003).

Impact: These studies permit the more rational design of gels for the separation of circular DNA. The gels can be designed or "tuned" for a particular analytical or preparative separation and thus give an improved separation of the circular DNA. The data and the apparatus can be used by researchers to prepare highly pure circular DNA in a few minutes.

Future Plans: We plan to use fluorescence microscopy to observe individual circular DNA molecules, tagged with fluorescent dyes, as they separate in gels. The motion of the DNA molecules as they move through the gels under the influence of an electric field will be measured and the data analyzed using models to describe the mechanism. We will collect data under a variety of gel conditions and from different DNA molecules to form a complete data set. This data will be analyzed to give us insight into the mechanism of separation and thus allow us to design improved separations.

Develop Imaging Tools to Characterize Fluorescence from Fluorophores Immobilized on Surfaces

Authors: Adolfus Gaigalas, K. D. Cole, L. Wang (831)

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: Imaging of fluorescent signals from surfaces has a number of important applications in biotechnology, including the detection of cell types, viruses, and measurement of target molecules. Advances in genomics and proteomics have made larger scale screening techniques possible. We are starting a program to study the effect of fluorescent probe structure and the surface environment on the measurement of fluorescence. We are studying the processes of photobleaching of fluorophores and fluorescent quenching. The research aims to improve the limits of detection of targets and to increase the quality of the data obtained.

Purpose: A large variety of new fluorescent reagents have been developed for the measurement of biological targets either in solution or on surfaces. These fluorescent dyes have to a large extent replaced the use of more dangerous and less convenient radioisotopes. However, the use of fluorescent reagents to detect biological targets on surfaces is not without their drawbacks, including the photobleaching of the fluorophore, fluorescence quenching of the light due to interactions in the microscopic environment on the surface, and background subtraction.

Major Accomplishments: We are measuring the process of photobleaching in solution and the effect of protective additives on the rates of bleaching. A fluorescent microscope has been set up to image targets including DNA molecules, viruses, bacteria, and microarrays. We are starting measurements of these targets and the evaluation of the images using commercial software programs. We are in the process of improving the microscope to improve the quality of the images obtained. These preliminary experiments have also indicated the importance of the sample preparation (labeling with the fluorophore) and the surface environment on the quality of the measurements obtained. Publications: Gaigalas, A. K.; Wang, L.; Vogt, R. F., "A frequency-domain technique for studying photodegradation process of fluorophores," Photochem. Photobiol. 76, 22-28 (2002); Gaigalas, A. K.; Wang, L.; Cole, K. D.; Humphries, E., "Photodegradation of fluorescein in solutions containing n-propyl gallate," in review.

Impact: Effective addictives help to prevent fluorophores from photodegrading and photobleaching. This serves to increase the limit of detection and the overall contrast to the auto-fluorescence background. In turn, this enhances and makes possible applications such as virus and bacteria counting and the imaging of individual microarray spots.

Future Plans: We plan to continue measurements of photobleaching for different fluorescent reagents and the effect of stabilizing additives. The effect of different surfaces and substrates on fluorescence quenching of fluorophores and background will be examined.

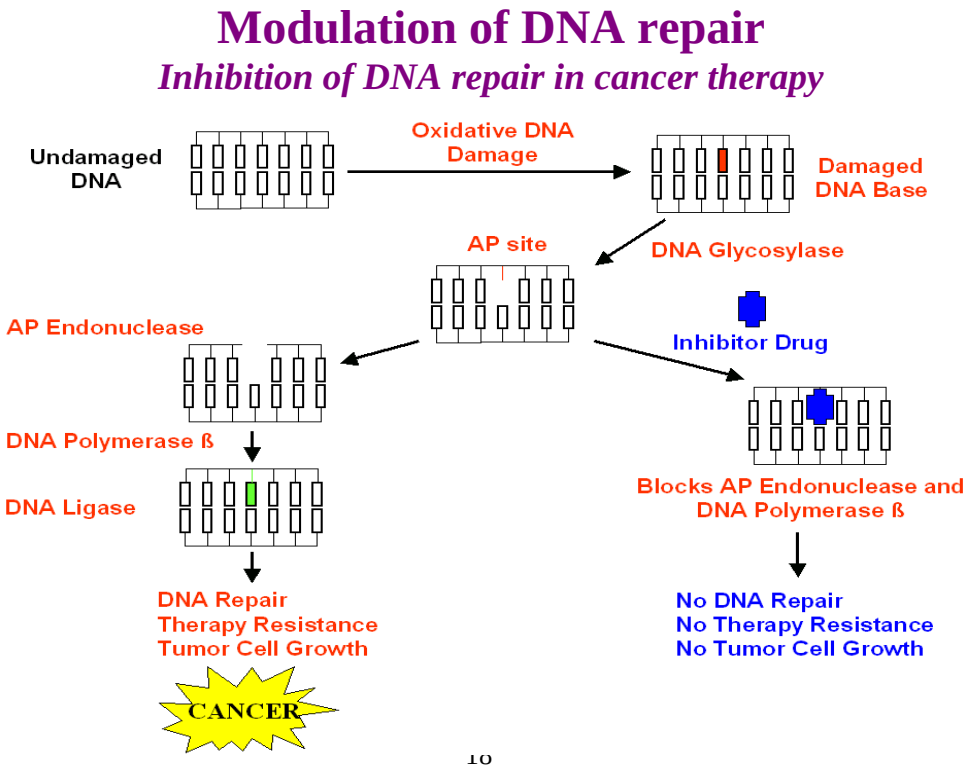
Substrate Specificities and Kinetics of Action of DNA Glycosylases Involved in Cellular Repair of Oxidative DNA Damage

Authors: *Miral Dizdar* (831)

CSTL Program: Health and Medical Technologies

Abstract: Oxidative DNA damage is generated by free radicals formed in living cells by normal metabolism and exogenous sources such as ionizing radiations and carcinogenic compounds. Numerous products are formed in DNA by oxidative damage. Most DNA products are cytotoxic and mutagenic with deleterious biological consequences. In living cells, repair systems exist to repair damaged DNA to sustain the genetic integrity and to prevent genomic instability that may lead to the onset and progression of cancer and other diseases. Deficiencies in DNA repair with elevated genomic instability are associated with a number of diseases. Oxidative DNA damage is mainly repaired by base-excision repair involving repair enzymes called DNA glycosylases. In our studies, we used the technique of gas chromatography/isotope-dilution mass spectrometry because of its ability to simultaneously measure multiple modified bases in damaged DNA. This technique enables precise determination of substrate specificities of DNA glycosylases by identifying as to which modified bases are or are not removed from damaged DNA by a given enzyme. Our studies showed that extensive variability exists in the substrate specificities of numerous DNA glycosylases.

These measurements lead to understanding of DNA repair mechanisms, thereby enabling methods to be developed for prevention of diseases and for new therapeutic approaches. This includes the improvement of cancer therapy by inhibiting DNA repair in drug- or radiation-resistant tumors by altering expression of DNA repair genes (which generate the DNA glycosylases and AP endonucleases shown in the figure). By coupling these altered expression patterns with DNA repair inhibitors that show strong selectivity in killing tumor cells in comparison to normal cells, new approaches to curing solid-tumor cancers are being developed.



Purpose: To develop methodologies to accurately measure the substrate specificities and kinetics of action of DNA repair enzymes.

Major Accomplishments: This work developed accurate methodologies to measure the repair of oxidative DNA damage and provided the first detailed studies regarding the substrate specificities and kinetics of action of DNA glycosylases. The paper reviewing these studies has been published in *Mutation Research*.

Impact: To provide researchers in industry and academia with accurate measurement technologies, this work supports the efforts to develop methodologies to elucidate mechanisms of enzymic repair of oxidative DNA damage.

Future Plans: These studies will be extended to the characterization of other DNA repair enzymes involved in base-excision repair of oxidative DNA damage.

A Fluorescence Assay Designed to Screen for Inhibitors of HIV-1 Nucleocapsid Chaperoned Maturation of the Dimerization Initiation Site

Author: **J. P. Marino (831)**

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: RNA-based drug discovery in the pharmaceutical industry requires general approaches for detecting and quantifying nucleic acid-protein interactions that can be used as high-throughput screens (HTS). As an extension of our studies of the dimerization and maturation of the HIV-1 Dimerizatoin Initiation Site (DIS) Stem-loop, we have developed a fluorescence assay to detect NCp7 catalyzed DIS maturation. Our assay uses a molecular beacon approach to detect hairpin strand exchange associated with NCp7 refolding of the DIS kissing dimer. One DIS hairpin is labeled with a fluorescent dye on the 5'-end and a the static quencher, dabcyl, on the 3'-end. Upon interaction with the second DIS hairpin to form the kissing dimer complex, the fluorescence emission of the dye remains highly quenched and static. In the absence of inhibitor, NCp7 catalyzed maturation of DIS kissing dimer, that is associated with strand exchange between the two hairpin stems, results in a burst of fluorescence from the reported dye as it no longer in close proximity to the quenching agent in the mature duplex form (Fig. 1). This fluorescence assay can form the basis for a HTS screen to identify novel inhibitors of the NCp7 mediated maturation of the DIS dimer.

Purpose: The development of specific inhibitors of retroviral protein-RNA complexes is of significant interest to the pharmaceutical industry since these complexes provide potential novel drug targets. A key roadblock to the realization of these goals is the availability of rapid and sensitive assays to measure and quantify the RNA interactions. The goal of this research was to develop a fluorescence-based HTS method suitable to screen for inhibitors that block a viral protein-RNA interaction.

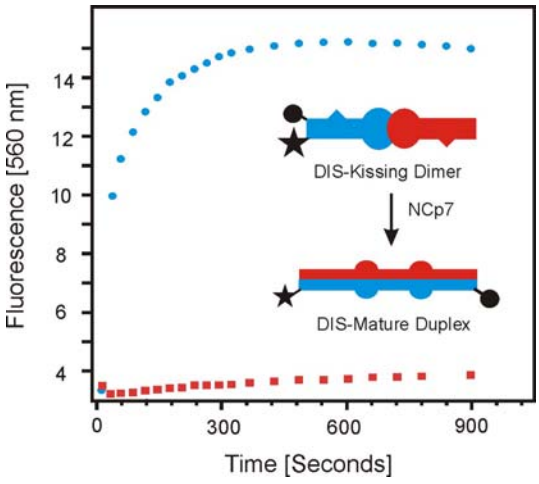


Fig. 1: Example of inhibition of NCp7 catalyzed DIS maturation by Neomycin at a concentration of 1 μM, as assayed using our molecular beacon detection approach. Fluorescence after addition of NCp7 to the DIS kissing dimer in the absence (blue) and presence (red) of Neomycin.

Major Accomplishments: A fluorescence method for rapidly screening and quantifying HIV-1 NCp7 chaperoned DIS maturation has been developed. This method has been validated as a screening tool using aminoglycoside antibiotics, like Neomycin, which have been shown to inhibit NCp7 activity.

Impact: Measurement technology developed in this study will have an impact on the biotechnology and pharmaceutical industries as well as the Structural Biology program at NIST. The research is also responsive to ATP's interest in projects directed at developing novel approaches for manipulating protein-nucleic acid interactions for possible therapeutic benefits or medical diagnostic purposes.

Future Plans: Through collaboration with researchers at the HIV-1 Drug Resistance Program at the National Institute of Cancer, we plan to apply our methods to screen large public domain libraries of low molecular weight compounds. Inhibitors of DIS and/or NCp7 identified in the HTS screen will then be optimized using rational design strategies based on NMR structural analysis.

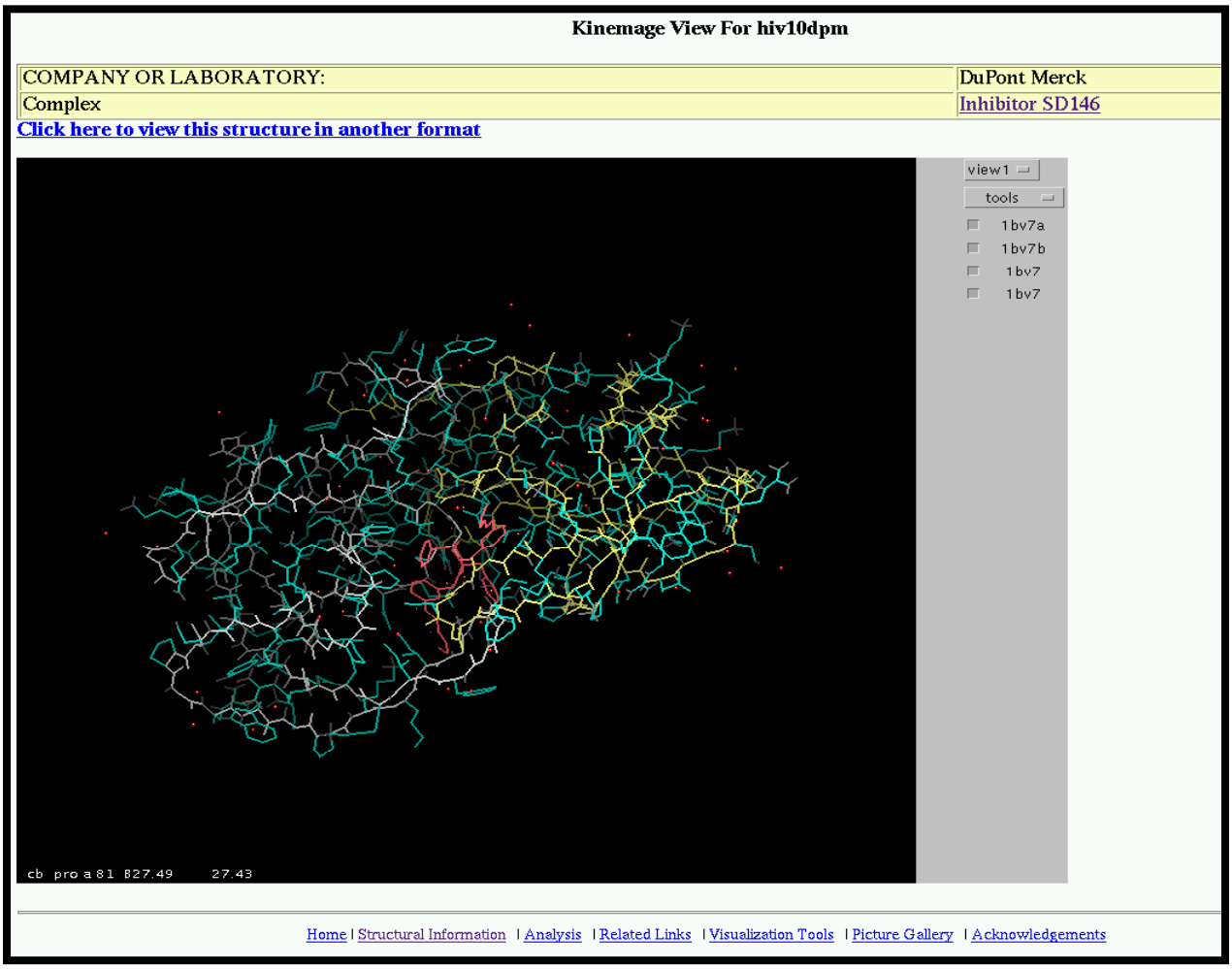
HIV Protease Structural Database

Authors: *T.N. Bhat, G.L. Gilliland, V. Ravichandran, M. D. Prasanna (831), and J. Vondrasek (NCI/NIH), A. Wlodawer (NCI/NIH)*

CSTL Program: Health and Medical Technologies

Abstract: The HIV Protease Database (HIVdb) is a web-based archive for HIV protease structural information. HIVdb contains experimentally determined three-dimensional structures of human immunodeficiency virus type 1 (HIV-1), human immunodeficiency virus type 2 (HIV-2), and simian immunodeficiency virus (SIV) proteases and their complexes with inhibitors or products of substrate cleavage. This resource provides macromolecular structural data on HIV and related proteases and tools to analyze this data to scientists engaged in basic research, in the development of drugs and in the development of therapies in the in the fight against AIDS. The structural holdings of the HIVdb are from two sources, the Protein Data Bank and from academic, government and industrial laboratories that have not deposited the coordinates in the PDB. The HIVdb can be accessed via the web at http://srdata.nist.gov/hivdb.

Purpose: To create and maintain an archival for the HIV Protease macromolecular data.



Major Accomplishments: In collaboration with the National Cancer Institute, HIV Protease Database has been developed to gather the structural information about the HIV-1, HIV-2 and SIV Proteases. The new web-based HIVdb permits the user to customize the query to retrieve the desired information. The query interface provides the option to select entries by the year in which the structures have been solved, resolution range, R Factor, strain in which the proteases were isolated from, inhibitor type, company or laboratory in which the structures were solved, virus type and citation or depositor author. In addition, the users have the option of dynamically structuring the report page that is generated. Our HIVdb is cited at Science's NetWatch section (Science, Vol 300, 1349, 2003), under "Databases: HIV's Achilles Heel".

Impact: This HIVdb's structural information offers new insight into the design of drug development towards manipulating HIV proteases.

Future Plans: The future plan is to convert web server language from Active Server pages (ASP) to Java based server languages, so as to give more option of incorporating readily available free software. Also, we are planning to include the option of structure based searching tool for the ligands bound to HIV proteases.

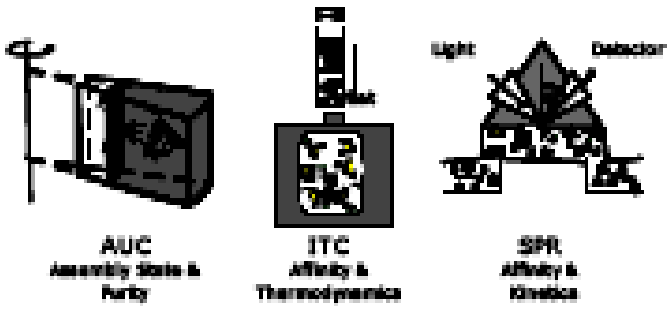
The MIRG 2002 Study: Assembly State, Thermodynamic, and Kinetic Analysis of an Enzyme/Inhibitor Interaction

Authors: *F. Schwarz (831), E. Eisenstein (831,UMBI), D. Myszka (U. of Utah), Y. Abdiche (U. of Utah), F. Arisaka (Tokoyo Institute of Tech.), O. Byron (U. of Glasgow), P. Hensley (Pfizer, Inc.), J. Thomson (Pfizer, Inc.), C. Lombardo (The Burnham Institute), W. Stafford (Boston Biom. Res. Institute) and M. Doyle (Bristol-Myers Squibb)*

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: Fully characterizing the interactions involving biomolecules requires information on the assembly state, affinity, kinetics, and thermodynamics associated with complex formation. The analytical technologies in the biotechnology industry often utilized to measure biomolecular interactions include analytical ultracentrifugation (AUC), isothermal titration calorimetry (ITC), and surface plasmon resonance (SPR). In order to evaluate the capabilities of industrial, academic, and government laboratories to implement these technologies, the Molecular Interactions Research Group (MIRG) of the Association of Biomolecular Resource Facilities (ABRF)

developed a standardized model system and distributed it to a panel of scientists employing AUC, ITC, and SPR.



The model system consisted of a well-characterized enzyme-inhibitor pair, namely bovine carbonic anhydrase II (CA II) and 4-carboxybenzenesulfonamide (CBS). Study participants were asked to measure one or more of the following: 1) the molecular mass, homogeneity, and assembly state of CAII by AUC; 2) the affinity and thermodynamics for complex formation by ITC; and 3) the affinity, kinetics, and thermodynamics of complex formation by SPR.

Purpose: The pharmaceutical industry has put increased effort into the discovery and development of protein-based therpeutics and reagents ("biologics") that exhibit highly specific interactions with their biological targets.. MIRG has recognized a need to develop standard reactions that can be used to validate the results from AUC, ITC, and SPR. Accordingly, samples of the CBS-CAII model system were distributed to a panel of scientists employing AUC, ITC, and SPR and the results of these measurements were evaluated.

Major Accomplishments: Samples were distributed to a total of 55 industrial, academic, and governmental laboratories that consisted of 7 AUC, 12 ITC, and 36 SPR laboratories. The results of the study showed excellent agreement to within 10 % on the binding affinities and within 2 % on the binding enthalpies between the ITC and SPR laboratories. The AUC results on the assembly state of the enzyme were in agreement with the ITC results on the stiochiometry of the reaction.

Impact: The Association of Biomolecular Resource Facilities that initiated this study is a non-profit standards organization sponsored by 33 companies and consisting of about 800 members from industry, academia, and government. The results of the study are posted on the www.ABRF.org website and will be published in the Journal of Biomolecular Techniques.

Future Plans: A similar study on a model protein-protein interacting system consisting of barnase and barstar mutants is presently under consideration.

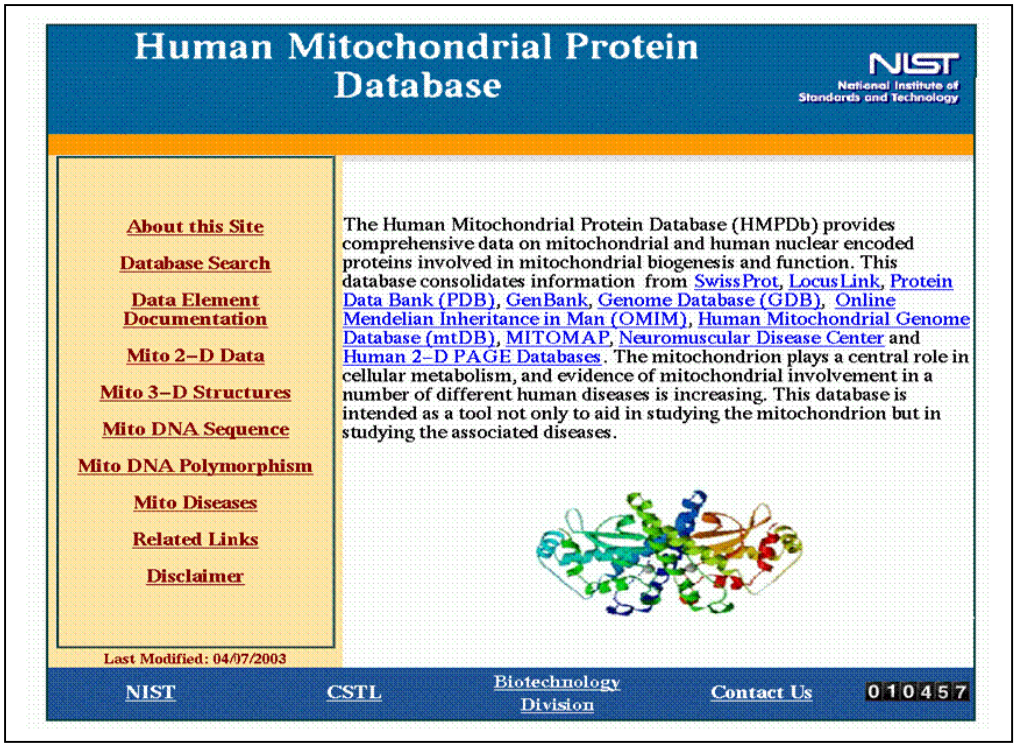
Human Mitochondrial Protein Database: A Resource for Human Mitochondrial Proteomics

Authors: *Veerasamy Ravichandran, Peter E Barker, Gregory B. Vasquez, T.N. Bhat, G.L. Gilliland (831), SJ Zull (ATP/NIST)*

CSTL Program: Health and Medical Technologies

Abstract: A publicly available web-based resource, the Human Mitochondrial Protein Database (HMPD), has been developed to unify data from a broad collection of related resources that provides information about human nuclear and mitochondrial-encoded proteins. This resource is designed to address the problem presented by the increasing amount, complexity and types of proteomics data available in public databases. At this point in time, no systematically established correlations between the different types of data have been established that allows meaningful comparisons. One of the primary reasons for this is the lack of data standards. Despite this, a wealth of data exists and is readily available from numerous web resources that is applicable to proteomics and have direct relevance to identifying proteins and assigning function. However, serious problems with the interactions between these resources arise because of problems with data exchange and interoperability. Although public proteomics data resources are highly informative, individually, the collection of available content would have more utility if provided in a standard and centralized context and indexed in a robust manner for a specialized area. HMPD currently consolidates information from many publicly available sources. The human mitochondrial protein database is available at

http://bioinfo.nist.gov:8080/examples/servlets/index.html



Purpose: The purpose is to develop a centralized database for mitochondrial proteins that will lead towards developing data standards for proteomics using the mitochondrion as a model system.

Major Accomplishments: A customizable interface has been developed to permit complex queries that include the name of the protein, tissue, mitochondrial compartment, chromosome number, molecular weight range, pI range, and keywords. Users can also restrict the data being searched to 2-D PAGE images, Locus links, the Genome Database, OMIM, or RefSeq information. To further narrow the search, selected results are grouped with gene name, Protein name and SwissProt ID, where the SwissProt ID is linked with detailed information for a protein of selection. A graphical tool was developed to visualize the human mitochondrial DNA sequences (16,569 bp). A mouseover feature has been added to include the annotated information. A polymorphism check tool has been implemented to dynamically compare the change in mtDNA changed between two populations. A standard reference data for two-dimensional gel electrophoresis has been implemented. Science's NetWatch has cited our web site under "RESOURCES: Probing the Cell's Powerhouse" (Science, Vol 300, 713, 2003).

Impact: There has been hundreds of thousands hit on our web site and as of October, 2003, about 16,000 researchers visited our web site.

Future Plans: Our future plan is to develop this web site as a standard reference data for mitochondrial proteomics experimental data from areas that impact proteomics will be archived and incorporated into the database (HMPDB). Data definitions and standard dictionaries will be developed to facilitate data acquisition, display and exchange.

Plasmid-based Standard for Quantitative Real Time Polymerase Chain Reaction (PCR) Measurements

CSTL Program: Food and Nutrition

Authors: *Marcia J. Holden, J. R. Blasic, Jr.(831)*

Abstract: Quantitative Real Time PCR is a powerful tool for the detection and quantitation of specific DNA sequences. It has broad application in basic scientific research, health care, pharmaceutical development, bio-warfare agent detection, and biotech crop measurements. In this methodology, specific target DNA is amplified by using the enzyme DNA polymerase. Throughout the PCR cycling procedures, the amplified DNA is monitored and measured by the use of fluorescent probes. While this is an extremely powerful tool, there are a large number of possibilities for introducing error in the measurement process. Thus, several factors need to be under proper control if one is to have robust and accurate measurement protocols as well as verifiable instruments and personnel performance. An important factor is the proper control and elimination of possible DNA contamination. Interestingly, there has been little work done on the comparability of the different instrumentation and detection platforms. We have addressed some of the measurement issues by designing a 200 base pair DNA sequence, which was used to prepare a plasmid suitable for PCR measurements. This material is being used for the validation of measurement protocols, equipment, personnel and varied laboratory practices applied to the measurement of specific DNA sequences. This effort involves 14 laboratories from around the world including the national measurement institutions of 11 countries. It is being coordinated by the Comité Consultatif pour la Quantité del Matière (CCQM) under the auspices of the Comité International des Poids et Mesures (CIPM).

Purpose: To investigate the utility of a synthetic DNA construct for the validation of measurement protocols, equipment, personnel and varied laboratory practices used in the quantitation of specific DNA sequences. This cross-platform material will provide information on the comparability of different platforms and detection probe types.

Major Accomplishments: The material for this study and future standard was designed from the ground up so that it could be used with different types of protocols, instrumentation and probes. It was also designed so that it resembled no known gene. The designed DNA sequence (200 base pairs) was synthesized by a self-assembly process and ligated into a plasmid. The plasmid was sequenced, propagated in bacteria, isolated and purified on cesium chloride gradients. PCR protocols were designed and assays conducted to verify the suitability of the material for this purpose. Studies were conducted to determine the recoverability and stability of the plasmid material after lyophilization.

A round-robin intercomparison pilot study was set up by the Comité Consultatif pour la Quantité del Matière (CCQM) under the auspices of the Comité International des Poids et Mesures (CIPM). The CCQM has the participation of 14 laboratories from around the world including national measurement institutions from 11 countries and the European Union. Multiple sets of lyophilized DNA materials were shipped to the participants. Test materials were prepared so that participants in the study could produce a standard curve of known DNA concentration and then analyze and report on the values of DNA samples of unknown concentrations. The

participants could select their instrumentation and probe type and were free to design whatever measurement protocol they deemed suitable. Several of the laboratories utilized more than one probe type and instrument. Very recently, the intercomparison participants returned their data and detailed information on how the studies were conducted. Analysis of the data is now underway.

Impact: The results of this study will provide basic information on the comparability of instrument and detection platforms for quantitative real time PCR. It will also shed light on sources of error and point to ways to improve and standardize methods. The study will lead to the improvement of the design of the material itself and clarify the utility of this type of material as a standard for monitoring the quality of quantitative PCR measurements.

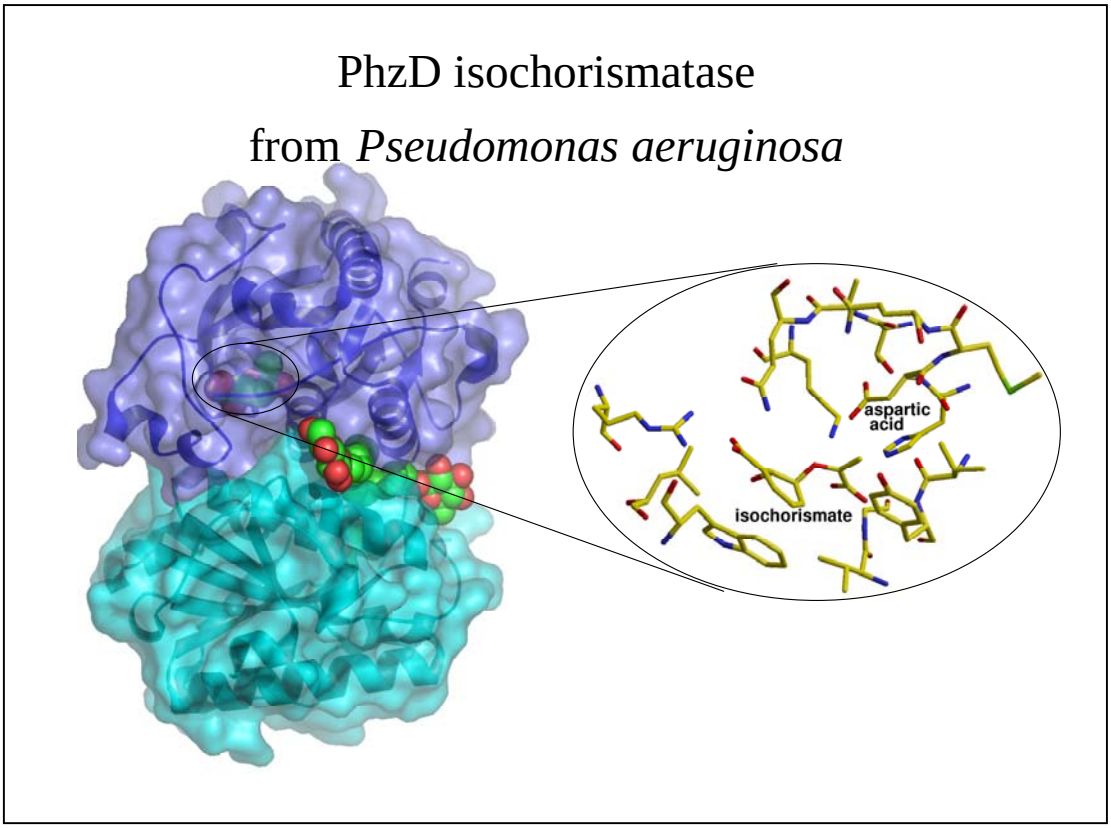
Future Plans: The analysis of the data from the first intercomparison study will point to future directions. It is likely that the plasmid material will be redesigned to accommodate detection platforms that have been recently developed.

Structure of Pseudomonas aeruginosa PhzD, an Isochorismatase from the Chorismate Biosynthetic Pathway

Authors : *Jane E Ladner(831), E. Eisenstein (831, UMBI), J. Parsons (UMBI)*

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: Industrial production of drugs, new biopolymers and indigo dyes, can be facilitated by increased understanding of the aromatic amino acid synthesis pathway. The increased ability to manipulate this pathway through metabolic and protein engineering will save non-renewable, petroleum-based feedstock chemicals. This study provides a description of the enzymes involved in the chorismate metabolic pathway by solving their three-dimensional structures, modeling the mechanisms of the chemical transformations, and mapping pathway control nodes involved in the biocatalytic conversion of glucose to aromatic hydrocarbons. Only when the detailed three-dimensional structures are known for the enzymes can the precise enzymatic mechanisms and relationships between the structure and the physical properties be predicted. Most recently, the structure of the product of gene phzD from *Pseudomonas aeruginosa* has been solved. The phenazines are biologically active (antibacterial, antifungal, antitumor) aromatic products synthesized mainly by *Pseudomonas* and *Streptomyces* as part of their chorismate pathways.



On the left, the structure of the protein phzD from the bacterium *Pseudomonas aeruginosa* is shown as a molecular surface with the secondary structure depicted as a ribbon inside the surface. Also buried inside the structure is the substrate for the enzyme shown as space-filling balls. On the right, the active site of the enzyme with the substrate, isochorismate, is show as a stick model. This enzyme is part of the pathway that synthesizes a bioreactive molecule that has antimicrobial properties. The labelled amino acid, aspartic acid, initiates the reaction by acting like a general acid in the protonation of the ether oxygen of the substrate. The determination of the crystal structure with the substrate provides the basis for hypothesizing how the enzyme works and thus can provide clues as to how we can control or alter its action.

Purpose: This work benefits pharmaceutical and chemical companies. The elucidation of natural biochemical pathways makes it easier to alter and utilize these pathways to make these and similar chemicals. In particular, the phenazines are difficult to synthesize but are important potential drugs.

Major accomplishments: The work this year has yielded a better understanding of the chorismate pathway. We have solved the structure of the product of gene phzD and are pursuing both biochemical and crystallographic methods to determine its role and substrate.

Impact: The synthesis of the aromatic biologically active molecules is a very important area of study because these pathways do not exist in animals. They exist solely in bacteria, fungi, and higher plants. The elucidation of these pathways can provide pharmaceutical researchers with a better understanding for the amplification or inhibition of the production of these molecules.

Future plans: We plan to continue our studies of the enzymes along the chorismate pathway. In particular, we are continuing to look at other enzymes in the phenazine pathway in *Pseudomonas* in order to more fully elucidate the mechanisms involved in the production of these biologically active products.

Title: DNA Repeat Element Mutation Frequency Measurements of the Polypurine/pyrmidine Tract Contained in the PKD1 Gene of Polycystic Kidney Disease.

Authors: *John P. Jakupciak. Catherine D. O'Connell (831); A. Bacolla (TAMU, TX); and R.D. Wells (TAMU, TX).*

CSTL Program: **Health and Medical Technologies**

Abstract: The presence of repeated sequences is a fundamental feature of genomes. Repeat elements are the simplest form of regularity and analyzing repeats gives first clues to discovering new biological phenomena in the same way as repeated words give a starting point to deciphering a script written in an unknown language. The human genome consists of repeating DNA patterns of various sizes, from very small to very large. Although the functions of these repeating regions are not well understood, they appear important for understanding the expression, regulation and evolution of DNA. In eukaryotic genomes, tandem repeats are involved in various regulation mechanisms that are still being discovered. Repeat elements are involved in human neurological disorders, such as fragile X syndrome. Instability of polypurine/pyrimidine repeats has also been shown to be associated with polycystic kidney disease (PKD).

Purpose: As part of the molecular measurements program at NIST, we sought to measure the mutation frequency of the polypurine/pyrimidine tract present in PKD and characterize the repeat element instability. This research is a second step in validating measurements associated with repeat expansions in human disease, supporting our past research on fragile X trinucleotide repeat elements. It is expected that the measurements derived from this research will assist in the standardization of repeat element measurements for both basic and clinical research needs.

Major Accomplishments: Our measurements support the concept that non B-DNA structures, formed by the repeat elements, are located at the mutation sites. Overall, the results indicate that human disease is associated with DNA mutations (break points) that coincide with non B-DNA conformations. This work provides the first detailed insight regarding DNA structural conformations and their function at DNA break points. The results of this study have been submitted to Science.

Impact: More than 10% of the human genome is composed of interspersed repetitive elements. Already more than a dozen human diseases have been associated with trinucleotide repeat elements. Therefore, identifying repeat elements that disrupt or contribute to gene dysfunction will significantly increase our knowledge of human disease. Our studies will provide fundamental measurements for researchers in industry and academia investigating DNA mutation(s) associated with human disease. This work supports the efforts to understand the function of repeat elements in the human genome. Characterization of repeat elements will provide insights into human predisposition and susceptibility for disease.

Future Plans: To develop a center of expertise in measuring DNA repeat elements. In collaboration with other government and academic laboratories, our current studies will be extended to annotate the entire human genome for repeating blocks of polypurine/pyrimidine elements.

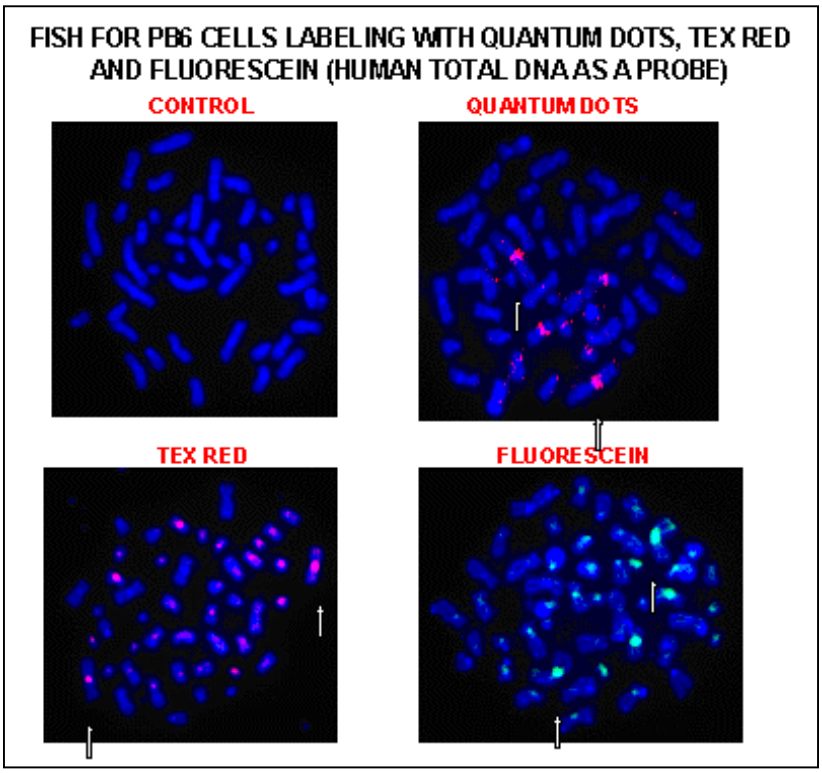
Semiconductor Nanocrystal Probes for Human Metaphase Chromosomes

Authors: *Peter E. Barker (831), Yan Xiao (831)*

CSTL Program: **Health and Medical Technologies**

Abstract: About one third of breast tumors have amplification of the gene for the tyrosine kinase receptor HER2/neu. In contrast to the other breast cancer patients, HER2 overexpressor tumors can be specifically targeted by the first rational design drug, Herceptin (trastuzumab), a humanized mouse monoclonal antibody that targets cells with high expression of the HER2 gene. Clinical testing for HER2 overexpressors follows two approaches, FISH (fluorescence in situ hybridization), a gene test measuring the HER2 gene at the DNA level, and an IHC (immunohistochemical) test that measures the amount of HER2 protein expressed on the cell surface. The IHC and FISH tests do not always select the same patients for trastuzumab therapy. In the process of developing a standard cell analyte with known HER2 levels, alternative fluorophores were investigated in a model system. Most FISH tests employ organic fluorescent tags. Recently, an inorganic tag of the binary compound CdSe, has become available. CdSe and other similar quantum dot labels are much more stable on exposure to light, and are brighter than conventional organic fluors. The CdSe labeled DNA probes were directly compared with fluorescein- and Texas Red-labeled probes, and proved to be superior for quantitative measurements of the type necessary for HER2 gene analysis. It was demonstrated that DNA probes used for FISH testing could be detected with quantum dot labels, and such a strategy may be more quantitative for medical testing than prior approaches.

Purpose: To measure DNA copy number by hybridization with fluorescent nanocrystal probes.



Specificity of fluorophore-streptavidin detection of biotinylated total human DNA in metaphase chromosomes.
Bar in panel C is 10 microns.
(A) Control (no fluorophore),
(B) streptavidin-Qdot 605 detection of chromosome 1q12 region (vertical and horizontal arrows),
(C) TexasRed-streptavidin detection of biotinylated DNA hybridized to 1q12 (vertical arrows)
(D) FITC-streptavidin detection of 1q12 sites (vertical arrows).

Major Accomplishments: This work demonstrates the physical advantages of quantum dots as labels for human metaphase chromosomes. The results of this study will be reported at the 2003 national meeting of the American Society of Human Genetics, and have been published in Nucleic Acids Research.

Impact: To provide researchers and clinical laboratories in industry and academia with alternative, stable and high luminosity fluorophores for medical and oncology testing.

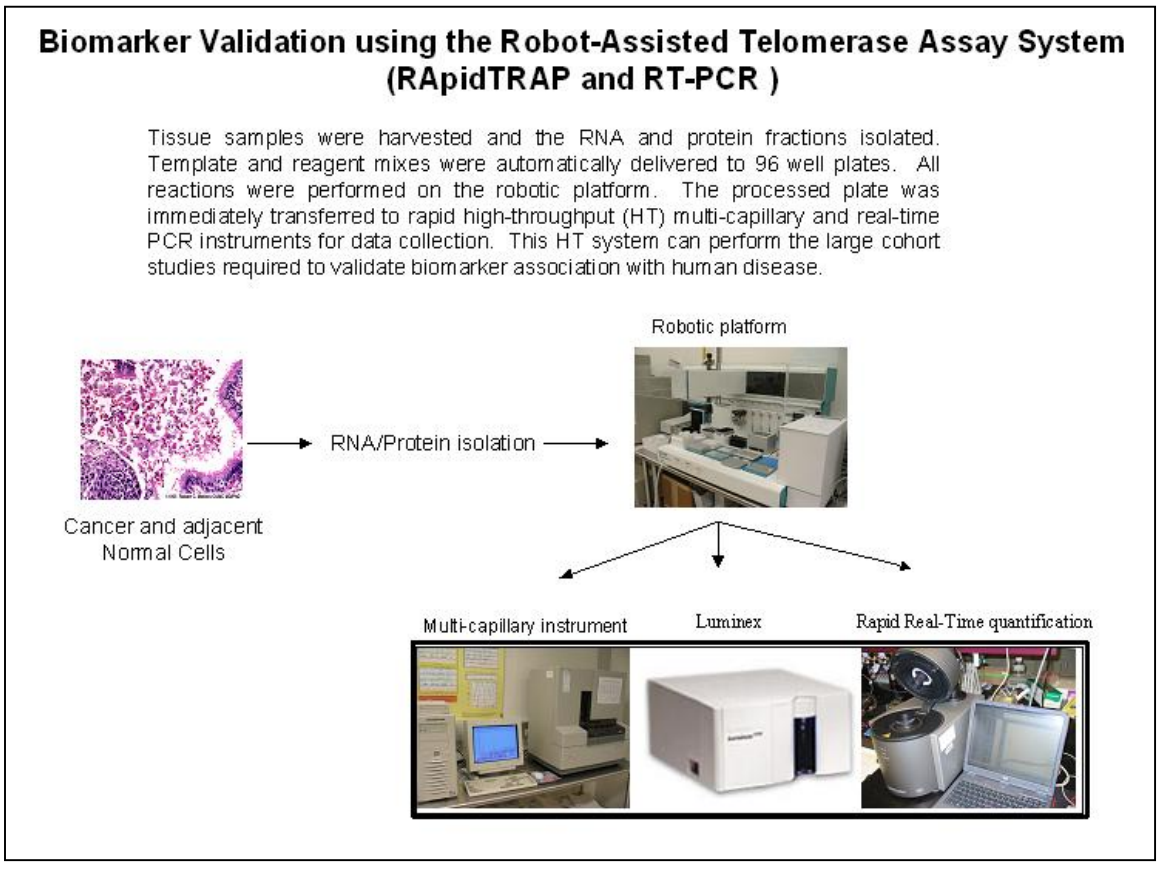
Future Plans: This study will be extended to the characterization of the gene HER2 for quantitative assessment of patients as candidates for trastuzumab therapy in breast cancer patients.

High-Throughput Analysis of Telomerase by Capillary Electrophoresis and Real-Time PCR

Authors: *Donald Atha (831), John Jakupciak (831) and W. Wang (NCI/NIH)*

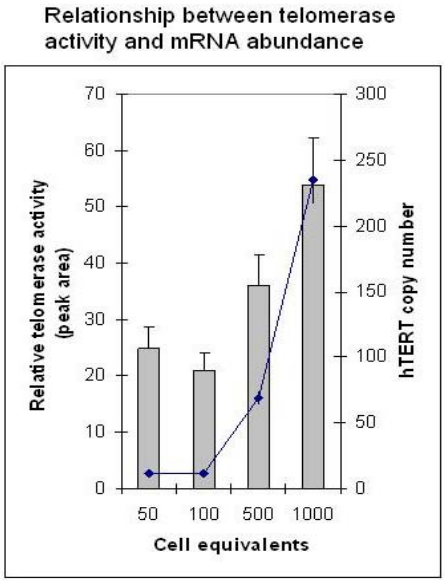
CSTL Program: Health and Medical Technologies

Abstract: Clinical studies show a direct association between telomerase activity and cells that are cancerous. The enzyme telomerase is expressed in 85-90% of all human cancers, but not in normal tissues. Currently, telomerase assays measure the small amount of enzyme activity in tissue lysates, urine or blood serum. The most commonly used system for the detection and quantification of telomerase activity (TA) is the PCR based assay known as the telomerase repeat amplification protocol or TRAP assay. Most of the TRAP assay systems use a slab-gel based electrophoresis system to size and semi-quantify the PCR products. We have improved the sensitivity and reproducibility of this system using capillary electrophoresis. However, such methods are still low-throughput. We have developed a high-throughput system or RApidTRAP (robot-assisted-TRAP). The MWG SE 4204 robot was used to generate TRAP/PCR samples for analysis using an ABI 3100 16-capillary array instrument. TA measurements were made using this method on (50 to 500) cells / reaction isolated from tumor cell extracts with a percent standard deviation (CV%) of 20. This was about half the variation obtained using the slab-gel based method. Measurement of hTERT mRNA was made using specific primers and probes on a LightCycler in the range of (10 to 7000) cells / reaction. The percent standard deviation (CV%) of these samples ranged from 0.21 to 4.47, which is a variation of about 2 to 40 RNA molecules per 1000. This is well within the criteria for differentiating between cancer and normal cells / tissue.



Purpose: To develop improved methods and standards for the analysis of telomerase. Clinical assays for telomerase in easily obtained body fluids would have great utility as non-invasive, cost-effective methods for the early detection of cancer. However, the use of telomerase as a biomarker has been hindered because of a lack of high-throughput methods essential for large cohort studies required for statistical evaluation. To this end we are developing high-throughput capillary electrophoresis and real-time PCR (gene expression) methods for the analysis of telomerase**.** This will result in a robotic sample handling system that will streamline validation of telomerase with increased sensitivity, reproducibility and automation over previous low-throughput methods.

Major Accomplishments: The most important result of this study is that the TRAP assay was successfully transferred from the manual to a robotic platform with results that were consistent in sensitivity and reproducibility. TA measurements were made using this method on (50 to 500) cells / reaction isolated from tumor cell extracts with a percent standard deviation (CV%) of 20. This was about half the variation obtained using the slab-gel based method. Measurement of hTERT mRNA was made using specific primers and probes on a LightCycler in the range of (10 to 7000) cells / reaction. The percent standard deviation (CV%) of these samples ranged from 0.21 to 4.47, which is a variation of about 2 to 40 RNA molecules per 1000. This is well within the criteria for differentiating between cancer and normal cells / tissue.



Comparison of hTERT mRNA and Telomerase Activity (TA). The relationship between the relative telomerase activity (shown in bars) and the amount of cellular hTERT mRNA. The amount of cellular hTERT mRNA (copy number) was measured in the A549 cells in comparison to the telomerase activity (total peak area) obtained from automated TRAP/PCR measurements

Impact: Since telomerase has been shown to be a primary candidate biomarker for cancer, improved measurement methods would impact numerous areas related to cancer diagnostics and would be used by customers in a wide range of areas that include academic and industrial research.

Future Plans: For analytical validation more extensive measurements will be made on patient samples obtained under IRB regulations. This combination of measurement of activity and mRNA expression will provide a more complete quantification of telomerase levels in tissues at various stages of activation.

Silicon Nanoparticles as DNA Labels

Authors: *Vytas Reipa and L. Wang (831)*

CSTL Program: Health and Medical Technologies and Exploratory Research

Abstract: To prevent photo-degradation and environmental sensitivity inherited by organic fluorophores, we explored the applicability of highly luminescent silicon nanoparticles as improved labels. Silicon nanoparticles (diameter = 1 to 2 nm) have been successfully conjugated to a 5'-amino modified oligonucleotide (60-mer) that contains a C6 linker between the amide and phosphate groups. The conjugation was implemented via two photo-induced reactions followed by a DNA labeling step through formation of a carboxamide bond. Photoluminescence of the conjugates is dominated by two blue bands (400 nm and 450 nm maximal) under 340 nm excitation. The quantum yield of oligonucleotide-conjugated nanoparticles was determined to be 0.08. This is the first report of the successful conjugation of silicon nanoparticles to an oligonucleotide in aqueous solution. The work provides a solid foundation for further investigations towards development of biological assays for disease-related research and diagnosis.

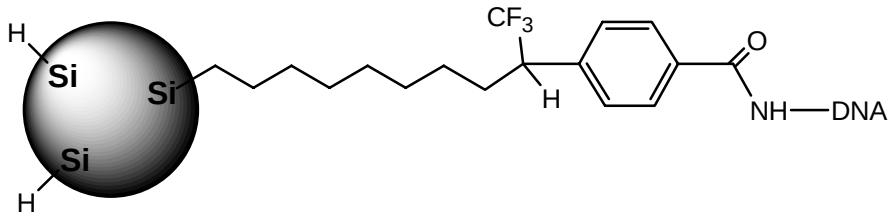
Purpose: To explore the applicability of highly luminescent silicon nanoparticles for DNA tagging in aqueous solutions. The silicon nanoparticles possess superior properties, such as high photo-stability and low toxicity, which are crucial for applications of luminescence (fluorescence)-based assays in clinical diagnosis.

Major Accomplishments: The multiple chemical steps for obtaining oligonucleotide-conjugated silicon nanoparticles (shown in the bottom Figure) include an electrochemical etching step and two photo-induced reaction steps followed by a DNA labeling step that allows the formation of a carboxamide bond between silicon nanoparticle and oligonucleotide. The procedure enhances aqueous solubility of hydrophobic silicon nanoparticles, which is a key requirement for the effective use of nanoparticles in biological applications. We successfully conjugated 1 to 2 nm diameter silicon nanoparticles to a 5'-amino modified oligonucleotide whose sequence is from rat tubulin genome. Under UV excitation, photoluminescence of the conjugates is dominated by two blue bands (400 nm and 450 nm maximal). The quantum yield of oligonucleotide-conjugated nanoparticles that were coated with a layer of 1-octene was determined to be 0.08 as measured against quinine sulfate in 0.1 M $HClO_4$ as a reference standard. We have demonstrated the feasibility of the present coupling chemistry to obtain silicon nanoparticle-labeled nucleotides.

Impact: The research will facilitate the development of alternatives to organic fluorophores. It is expected that the alternatives will be superior in the following respects: high luminescence quantum yield, high photo-stability, low toxicity, and size-dependent luminescence. These properties are critically important for biological applications such as DNA sequencing, microarrays for high-throughput screening, flow cytometry measurements, and fluorescence in-situ hybridization.

Future Plans: Spectral characterizations of silicon nanoparticle-DNA conjugates suggest moderate size inhomogeneity of the particles; therefore size exclusion chromatography separation would be justified in addition to ultracentrifugation prior to 1-alkene monolayer binding. Moreover, the use of a longer carbon chain would be expected to provide a better-ordered self-assembled monolayer (SAM), thus lowering non-radiative recombination channel

concentration and increasing photoluminescence quantum yield. These two improvements will likely enhance the luminescence quantum efficiency of the nanoparticles, which is critical for applications such as fluorescence in-situ hybridization in chromosomes.



Evaluation and Automation of Single Nucleotide Polymorphism Typing (SNP) Methodologies for Mitochondrial DNA (mtDNA).

Authors: *Margaret C. Kline, Janette W. Redman, Peter M. Vallone, and John M. Butler (831)*

CSTL Program: Forensics and Homeland Security

Abstract: Mitochondrial DNA (mtDNA) is used in human identity testing for mass disaster reconstruction, missing persons investigations, and forensic casework involving highly degraded DNA samples. Procedures for analysis of mtDNA are laborious and expensive because they typically involve DNA sequencing of more than 600 base pairs in the control region across hypervariable region I (HVI) and hypervariable region II (HVII). Techniques for rapid screening of samples, particularly to weed out ones that do not match, will enable more samples to be run by forensic DNA laboratories in a timely and cost-effective manner.

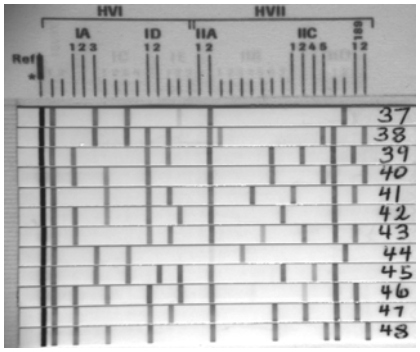
Roche Molecular Systems has developed a mtDNA LINEAR ARRAY system that probes ten sites in HVI and HVII. Their manual process of hybridization and detection was adapted to an automated format at NIST as part of beta-testing their product. In addition, we ran 666 U.S. population samples to evaluate the power of discrimination for the mtDNA LINEAR ARRAY test. The 666 NIST samples resulted in 282 different types, of which 185 were seen only once in our population. These unique samples represent 65.6% of the types seen and 27.8% of the people. Several groupings of samples seem to represent specific ethnic groups. Fifty-one samples were determined to be mitotype 1111111111 by LINEAR ARRAYs. After SNP typing the 51 samples could be sub-typed to 12 different SNP types with four samples being unique.

Purpose: Beta-testing of HVI/HVII mtDNA LINEAR ARRAYs (Roche Applied Sciences,Indianapolis, IN), and automation of the process.The LINEAR ARRAYS offer forensic laboratories doing human identification a rapid test for screening samples that may or may not require labor-intensive full sequencing.

Major Accomplishment: Development of an automated procedure for processing the mtDNA LINEAR ARRAYS using a Tecan / Profiblot hybridization robot. Analysis of the mtDNA mitotypes / haplotypes in 128 Hispanic, 252 African American, and 286 Caucasian population samples, which doubled the available population data on U.S. samples. Discovered that the pH of the washing buffer is critical to obtaining successful results—information that is now included in the technical manual for this product.

Impact: LINEAR ARRAYS allow laboratories without labor-intensive sequencing capabilities to rapidly screen mtDNA samples.

Future Plans: A paper describing the automated LINEAR ARRAY processing is in process. Future work with developing new SNP assays to increase the power of discrimination for mtDNA.



This figure shows the mtDNA LINEAR ARRAY results for 12 different DNA samples. These samples were processed in an automated fashion in a method developed at NIST as part of beta-testing this Roche Molecular Systems product.

Creation and Characterization of SRM 2395—A Human Y-Chromosome DNA Profiling Standard

Authors: *John M. Butler, Richard Schoske, Peter M. Vallone, Janette W. Redman, Margaret C. Kline (831)*

CSTL Program: Forensics and Homeland Security

Abstract: The Y-chromosome is passed largely unchanged from father to son and contains genes that endow the recipient with male characteristics. Forensic DNA testing laboratories, genetic genealogy companies, and academic groups studying human history perform Y-chromosome testing by examining a number of genetic markers. The ability to obtain consistent and reliable results between laboratories that share information through DNA databases is critical. To address this need, NIST has developed a human Y-chromosome DNA profiling standard, SRM 2395, that can be used to check the accuracy of analytical methods used to measure Y chromosome variation at common short tandem repeats (Y-STR) and single nucleotide polymorphism (Y-SNP) markers. One unit of SRM 2395 consists of six frozen tubes of DNA, each of which contains approximately 100 ng of well-characterized extracted human genomic DNA. There are five male samples (Components A-E) and one female sample (Component F) packaged in a single box. The female specimen may be used as a negative control for Y-chromosome specific assays. The Certificate of Analysis lists typing results for 31 Y-STR and 42 Y-SNP markers commonly used in Y-chromosome testing. DNA sequencing has been performed on 22 of the 31 Y-STRs to permit an even more extensive characterization of those markers.

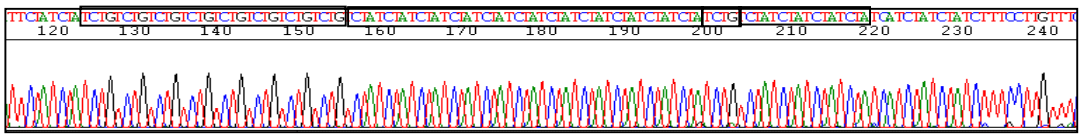
Purpose: Produce a Standard Reference Material with multiple male DNA samples that have been characterized at common Y-chromosome short tandem repeat and single nucleotide polymorphism markers to enable calibration of results obtained with either research assays or commercial kits.

Major Accomplishment: Characterization of five male DNA samples (components A-E) with genetic markers contained in all current commercial Y-STR and Y-SNP kits. In addition, 22 of the 31 Y-STRs have been sequenced across all five male components, which will ensure accurate typing can be performed with any PCR primer set used in the future.

Impact: SRM 2395 is being used already in genetic genealogy companies such as Relative Genetics to calibrate their results. New Y-STR typing kits from ReliaGene Technologies and Promega Corporation have used our SRM to verify the accuracy of their assay. These commercial kits are gaining acceptance in forensic DNA laboratories and will be used for many years to come to analyze sexual assault evidence.

Future Plans: A paper is in preparation to describe the work done in characterizing the SRM components. We plan to update the male SRM components with extra Y-STR and Y-SNP marker information as additional markers come into common use in Y-chromosome testing.

Figure shows the sequencing results for the Y-STR marker DYS390 on SRM 2395 component E. The sequence indicates that there are 24 repeats present with the motif of $[TCTG]_8$ $[TCTA]_{11}$ $[TCTG]_1$ $[TCTA]_4$



Structure to Function: Using a Structural Genomics Approach

Authors: *Gary L. Gilliland (831), Jane Ladner (831), Alexey Teplyakov (UMBI), Galina Obmolova (UMBI), Osnat Herzberg (UMBI), John Orban (UMBI), Andrew J. Howard (Illinois Institute of Technology), and John Moult (UMBI)*

CSTL Program: Pharmaceutical and Biomanufacturing

Abstract: An effort involving the large-scale structure determination of hypothetical (having no known structure or sequence homology with proteins with known functions) proteins from Haemopilus influenzae or homologs from other organisms has been underway to assess how structural information can influence functional assignment. The structures of more than 40 proteins have been determined providing new insight into the function of these proteins, identifying a number of new protein folds, and identifying a number of potential targets for drug development. The future effort will focus on applying this technology to investigating the structural implications resulting from the process of alternative splicing in higher organisms.

Purpose: The three-dimensional structures of Haemophilus influenzae proteins and homologous proteins from other organisms whose biological functions and three-dimensional structures are unknown are being determined as part of a structural genomics project to assess whether structural information can assist in assigning the functions of proteins. The structures of the hypothetical proteins are being used to guide further studies and narrow the field of such studies for ultimately determining protein function. In addition to aiding in functional assignment, this effort is identifying new protein folds and a number of possible new targets for drug development.

Major accomplishments: With large-scale genomic sequencing efforts meeting with success, post-genomics research efforts are focusing on functional assignment of gene products. One approach for the functional assignment of hypothetical proteins is the use of structural data determined by crystallographic or NMR methods.(1) More than forty structures have been elucidated or are in the process of being determined. A summary of the efforts is available at http://s2f.carb.nist.gov/. Three examples of these structures include the YcdX, YchF, and YajQ proteins. The YcdX protein from Escherichia coli was cloned, expressed, and the crystal structure determined at 1.6-Å resolution.2 YcdX has an unusual topology of a b7a7 barrel as compared to the more common b8a8 (TIM) barrel (Figure 1). All -strands are parallel, and their order is consecutive. The deep cleft at the C-terminal side of the barrel contains three metal binding sites ligated to the imidazole and carboxylate groups of the protein. Only four proteins with known structures have a trinuclear zinc catalytic site. All four (nuclease P1, endonuclease IV, alkaline phosphatase, and phospholipase C) hydrolyze the phosphoester bond. This suggests a similar activity for YcdX. The YchF protein from Haemophilus influenzae was cloned, expressed, and the crystal structure determined at 2.4-Å resolution.3 The polypeptide chain is folded into three domains The N-terminal domain has a mononucleotide binding fold typical for the P-loop NTPases. An all a-helical domain consisting of two long coiled-coils and one short a-helix is formed by an 80-residue insertion after the central b-strand of the N-terminal domain. The C-terminal domain features a six-stranded b-sheet forming a half-barrel that curves around an a-helix. The crab-like three-domain structure of YchF suggests the binding site for double-stranded nucleic acid located in the cleft between the all a-helical and C-terminal domains. Fluorescence measurements have demonstrated the ability of YchF to bind a nucleic acid duplex, and have also shown the specific binding of GTP. Taken together with other experimental data and given the universal phyletic distribution of this family of proteins, these

results suggest that YchF may be part of a ribonucleoprotein complex and may function as a GTP-dependent translation factor. The YajQ protein was cloned, expressed, and the crystal structure determined at 2.1-Å resolution by applying the multiwavelength anomalous dispersion method to a mercury derivative. The polypeptide chain folds into two domains with identical topology. Each domain has a four-stranded antiparallel b-sheet flanked on one side by two a-helices. This structural motif is a characteristic feature of many RNA-binding proteins. The tetrameric structure observed in the crystal suggests a possibility of binding two stretches of double-stranded nucleic acid. A possibility may be considered that YajQ binds a bacterial analogue of the eukaryotic SH3 domain implicating its function in possible regulatory protein-protein interactions. Whether YajQ binds nucleic acids, or interacts with the SH3 domain, or combines both activities, remains to be established in the future experiments, which will shed more light on the molecular function of YajQ.

Impact: This project involving a structural genomics approach has demonstrated that working on a number of proteins in parallel, chosen on the basis of a unifying scientific theme, is a powerful strategy. A large body of structural information has been obtained for hypothetical bacterial proteins while methods development to accomplish large-scale structure determination were on going. The structures, in addition to aiding in functional assignment, which must be followed up by further biochemistry and biophysics studies, this effort is identifying new protein folds and a number of possible new targets for drug development.

Future Plans: The approach developed in this project will focus next on alternative splicing, a new and important scientific discovery, resulting from the analysis of the genomic scale sequence information on higher eukaryotes. The effort will take advantage of the infrastructure that has been developed for the structural genomics effort reported here, with changes introduced for a focus on eukaryotic proteins, and improvements based on our experience so far. How does the structural mediation of functional diversity resulting from alternative splicing compare with that resulting from more traditional gene duplication and divergence of protein function? What are the implications for the evolution of new protein folds? What contribution does alternative splicing make to the great functional complexity of higher eukaryotes? The project will address these questions through structural studies. To this end, we will clone, express, purify and obtain the structures of a set of alternatively spliced proteins.

Enzyme-Catalyzed Synthesis of Optically Active Drug Intermediate Compounds Using Supercritical Fluids as the Solvent

Authors: *Y. B. Tewari (831), T. Ihara and K. W. Phinney (839)*

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: Enzyme-catalyzed reactions are of interest to the food and pharmaceutical industries. In the pharmaceutical industry, these reactions are used for the synthesis of chiral drug intermediates. Because these reactions are generally carried out in organic solvents, the generation of large volumes of organic waste is of concern. Recently, supercritical carbon dioxide ($SCCO_2$) has been used as an alternative solvent for biochemical reactions and has been shown to have several advantages over commonly used organic solvents. Specifically, carbon dioxide is non-toxic, non-flammable, available in abundance, and is environmentally friendly and recyclable. The low viscosity and high diffusivity of $SCCO_2$ serve to provide favorable mass transfer properties. In this work, a reaction vessel was designed, fabricated, and tested to study enzyme-catalyzed reactions in $SCCO_2$ media. The equilibrium constants of the lipase-catalyzed transesterification reaction (benzyl alcohol + butyl acetate = benzyl acetate + 1-butanol) were studied in $SCCO_2$. The equilibrium constant of this reaction in $SCCO_2$ is similar to those obtained in *n*-hexane, toluene, and other non-aqueous solvents. These results also show that the rate of reaction in $SCCO_2$ media is higher than those in organic solvents.

Purpose: To develop a method using supercritical carbon dioxide as the solvent for the enzyme-catalyzed synthesis of optically active drug intermediates.

Major Accomplishments: This is the first equilibrium study that has been performed in supercritical carbon dioxide. We have demonstrated that supercritical carbon dioxide provides kinetic and environmentally friendly advantages for carrying out enzyme-catalyzed reactions. Publications: Tewari, Y. B.; Ihara, T.; Phinney, K. W.; Mayhew, M. P., "A Thermodynamic and Time Course Study of the Lipase-Catalyzed Transesterification of Benzyl Alcohol and Butyl Acetate in Supercritical Carbon Dioxide Media," submitted to *J. Mol. Catal. B: Enzymatic*.

Impact: This work provides comparisons between enzyme-catalyzed reactions in traditional non-aqueous solvent systems and in supercritical carbon dioxide. The data will facilitate implementation of supercritical carbon dioxide as a solvent for biocatalysis in the biotechnology and pharmaceutical industries. Alternatives to organic solvents are important for reducing the levels of residual solvents in pharmaceutical products and for reducing the environmental burden resulting from solvent waste.

Future Plans: Studies of additional enzyme-catalyzed reactions in supercritical carbon dioxide are planned. These investigations will focus on reaction systems of industrial significance and relevance.

Theoretical Protocols for Calculating Structural, Spectroscopic, and Reaction Properties of Enzyme Active Sites

Author: *Morris Krauss (831); Yong S. Lee (NIH, Bethesda, MD)*

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: The reaction path and the electronic structure of enzyme active sites are analyzed to explore the common electronic character of enzyme catalysis. Applications are to a variety of enzyme types chosen for their importance in validating the methodology and for applications to industrial, military, and medical problems. The methodology can be applied to the same or analogous enzymes within families of organisms or between different families to discern catalytic motifs and functional properties of active site residues. The very large size of the enzyme systems requires a means of dividing them into a reactive region treated quantum mechanically and the protein and solvent environment that is treated classically. Two flavors of coupling the quantum region to the classical are used in this research. In one, the quantum region interacts with effective fragment potentials (EFP) that represent the electrostatic, polarization, and repulsive interaction. The EFP are fitted to the quantum interactions for model systems and produces a consistent interaction across the quantum/classical interface. Classical and quantum calculations that determine the qualitatively correct reaction path are initiated using crystal structures with bound analog inhibitors. The reaction path can then be used for virtual screening of inhibitors for drug development. The ability to calculate the substrate reaction path for homologous enzymes from different types of cell organisms and simple mutants provides a flexibility and directness in the theoretical virtual screen that should make it very competitive with most experimental programs.

Purpose: Both the QM/MM (Quantum Mechanics/Molecular Mechanics) and EFP codes are part of a protocol for studying enzyme behavior. Crystal structures with bound analog inhibitors are used to initiate classical and quantum calculations that determine the qualitatively correct reaction path. With the reaction path in hand, optimization and dynamics of the entire protein/ligand system with ab initio QM/MM can be considered. Also, spectroscopic properties are calculated from both the EFP and QM/MM data

Major Accomplishments: In bacteriorhodopsin, proton transfer from the cyctoplasm to the extracellular side is initiated from protonated asp96 in the cytoplasmic region towards the deprotonated Schiff base. This occurs in the transition from the late M photocycle state to the N state. A quantum mechanics/molecular mechanics (QM/MM) model is constructed from the bacteriorhodopsin E204Q mutant crystal structure. Structural transformation occurs in the M state allowing a channel to form between the asp96 and Schiff base that permits a water chain between these groups. Since a part of this channel is lined with hydrophobic residues, there has been a question on the mechanism of proton transfer in a hydrophobic channel. Ab initio dynamics using the CHARMM/GAMESS methodology was used to calculate the transfer of the proton through a hydrophobic channel, a new concept that is described in a paper submitted to the Journal of the American Chemical Society. Using the energy optimized structures and the molecular dynamics trajectories, the entire vibrational spectrum of bacteriorhodopsin has been calculated to compare with experimental observations.

Impact: Determination of a qualitatively correct reaction path is required before a QM/MM calculation can be undertake. The structures of the reaction path in the enzyme active site can be used to design high-throughput screens for assessing reactant, transition state, and product analogues for lead compounds for drugs. The ability to generally calculate vibrational spectra for

hydrogen bond networks bound in proteins will provide an important tool for analyzing the function of these systems.

Future Plans: Now that the reaction path for the two-zinc enzyme mechanism of zinc lactamase has been published, this work can be extended to other antibiotic substrates with different ionicities and to a comparison with other lactamases.

DNA Damage by the Antitumor Drug 3-Amino-1,2,4-benzotriazine 1,4-dioxide (Tirapazamine)

Authors: *M. Birincioglu, H. Rodriguez, M. Dizdar (831); P. Jaruga (UMBC, MD); and G. Chowdhury and K. Gates (University of Missouri)*

CSTL Program: Health and Medical Technologies

Abstract: Hypoxia is a typical feature of cancerous tumors. Drugs activated in hypoxia have wide potential applications, particularly in overcoming the resistance of hypoxic tumor cells to chemotherapy. Key features of such drugs include redox properties activated by reductase enzymes such as flavoproteins. In hypoxic tumor cells, the enzymatically reduced drug and/or its intermediates lead to formation of reactive hydroxyl radicals that effectively damage DNA of tumor cells, resulting in destruction of tumors. Tirapazamine is such a bioreductively activated DNA-damaging agent that selectively kills the hypoxic cells found in solid tumors. This compound shows great clinical promise and is currently being examined in a variety of clinical trials, including several phase III studies. It is well established that DNA is an important cellular target for tirapazamine; however, the structural nature of the DNA damage inflicted by this drug remains poorly understood. As part of an effort to understand the chemical events responsible for the hypoxia-selective cytotoxicity of tirapazamine, the studies reported here were designed to characterize this drug's ability to damage DNA. Here, we used gas chromatography/mass spectrometry and liquid chromatography/mass spectrometry to characterize and quantify oxidative DNA damage mediated by tirapazamine. A multiplicity of modified DNA bases were identified and quantified. It appears that the hypoxic conditions, under which tirapazamine operates, along with the unique chemical properties of the drug, yield a unique variety of DNA damage that includes poorly repaired, potentially cytotoxic DNA base lesions that block DNA transcription and replication. Overall, the results indicate that DNA base damage may contribute to the biological effects of tirapazamine *in vivo*.

Purpose: To identify and measure oxidative DNA damage generated by tirapazamine under hypoxic conditions.

Major Accomplishments: This work provides the first detailed insight regarding the structural identity and quantities of the DNA lesions caused by this new anticancer drug. The results of this study have been published in Journal of the American Chemical Society.

Impact: To provide researchers in industry and academia with expertise in the measurement of oxidative DNA damage, this work supports the efforts to develop methodologies to test new anticancer drugs similar in chemical structure to tirapazamine and to help elucidate the mechanism of action of hypoxia-selective antitumor drugs and their efficacy to destroy tumor cells, but not normal cells.

Future Plans: This study will be extended to the characterization of DNA damage by this drug *in vivo*, and by other structurally similar hypoxia-selective antitumor drugs.

Predicting Vibrational Spectra of Dissolved Salts

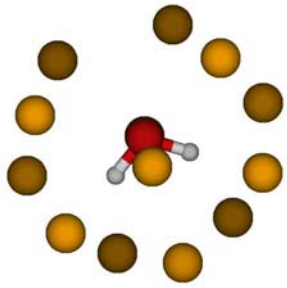
Authors: *LiLi Wang (831), K. K. Irikura (838)*

CSTL Program: Pharmaceuticals and Biomanufacturing

Abstract: Vibrational spectroscopy is an important technique in analytical chemistry. Compounds are identified by comparing observed spectra with those of candidate compounds, which must be previously known or predicted by theory. Despite the ubiquity of ions throughout chemistry, there are no reliable theoretical procedures for predicting their vibrational spectra in solution. The goal of this project is to produce a reliable, practical procedure to fill this need. This would have widespread impact for the analysis of organic and inorganic compounds in pharmaceuticals and manufacturing. To date, the most popular, cheapest models have been found useless, and better models are being investigated.

Purpose: Aqueous ions are commonplace and important in biology, geochemistry, many manufacturing processes, and elsewhere. Vibrational spectroscopy (infrared, Raman, and near-infrared) is often used to analyze or monitor such systems. For example, most biological molecules are ions, and vibrational spectroscopy is important for identifying their structures and interactions. For chemical analysis, molecules can only be identified by spectroscopy if the spectra of candidate molecules can be predicted in some way, for comparison. Modern theoretical methods can now predict gas-phase spectra with useful precision, but predicting condensed-phase spectra is more challenging. Ions are especially difficult because they interact strongly with the solvent or matrix, often causing dramatic changes in the spectrum. Although significant progress has been achieved in modeling solvation, the primary emphasis has been on energetics, with a secondary emphasis on molecular structure. The performance of solvation models in predicting vibrational spectra using quantum chemistry has received little attention.

Major Accomplishments: The inexpensive continuum models tested have failed. Consequently, computations with explicit solvent molecules are now underway. Two reasonable electronic-structure models have been identified and compared for the prototype ion H_2O^+ solvated by neon, which was chosen as the initial, simple case for which benchmark experimental data are available. Increasingly large computations still have not reached the bulk limit. A sample structure for $H_2O^+(Ne)_{13}$ is illustrated in the bottom Figure.



Impact: Reliable theory will permit observed spectra to be assigned with confidence, thus providing the necessary interpretation of the experimental data. The current limited success is essential for aqueous frequency predictions.

Future Plans: Ion-solvent clusters are mechanically “soft”, with multiple, competing structures. The corresponding effect upon the vibrational spectrum remains to be investigated. Simulated-annealing approaches will also be applied to find the most stable structure for each cluster. Both procedures will help in reaching the bulk limit, a major goal of this project.

Typing Single Nucleotide Polymorphisms Located in the Y Chromosome

Authors: *Peter M. Vallone and John M. Butler (831)*

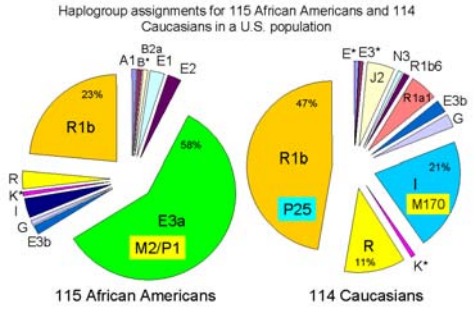
CSTL Program: Forensics and Homeland Security

Abstract: A single nucleotide polymorphism or SNP is a single base change in the human genome. A SNP is predicted to occur every 1000 bases or so in the human genome. There is interest in typing SNPs for human identity (forensic) purposes. More specifically, SNPs located on the Y chromosome are attractive typing candidates due to their low mutation rate and male specificity. The fact that the Y chromosome is only found in males makes genetic markers located upon it useful in mixed male-female sample (as found in a sexual assault crime). The low mutation rate of the Y chromosome SNPs implies potential for determining the geographical origin of an individual. To date, few studies have been performed evaluating the frequency of known Y chromosome SNPs in U.S. populations. By evaluating the frequency of 50 well-characterized SNP sites in over 200 U.S. population samples we provide a useful database of DNA typing information.

Here, we used a commercial DNA typing kit and custom assays developed in-house to determine the frequency of 50 Y SNPs in 2 U.S. populations (African American and Caucasian). The two assay methodologies, hybridization and primer extension, indicated full concordance for typing 8 overlapping SNP sites. The overall success rate for typing that samples was over 99% for either methodology. A pie chart summary of the data is illustrated in the figure. SNP frequency distributions for the 50 genetics markers typed indicated groupings that were unique to each population. 47% and 23% of the Caucasian and African American samples fell into haplogroup R1b respectively. 58% of the African American samples were found to be in the haplogroup E3a. It is of interest to note that none of the Caucasian samples fell into haplogroup E3a making it a population specific grouping. This information is useful in determining the forensic utility of using Y-SNP in forensic investigations.

Purpose: Development of multiplex assays for the evaluation of Y chromosome SNPs in a human identification context.

Major Accomplishments: We have completed one of the first studies involving Y-SNPs frequencies in U.S. populations. The results of this study have been submitted for publication to the Journal of Forensic Science. An oral presentation describing this work was presented in Arcachon, France at the International Society of Forensic Genetics (September, 2003).



This figure represents the distribution of various Y chromosome haplogroups (Hg) for 2 U.S. populations. Hg R1b represents the majority of the Caucasian samples while 58% of the African American samples are in Hg E3a. Note there is an admixture of 23% of Hg R1b (major Caucasian Hg) observed in the U.S. African American population.

Impact: Providing researchers in the forensic community with population data on newly discovered genetic markers. This work supports companies who have an interest in developing new forensic testing kits. By providing typing data for the 50 Y-SNPs in a U.S. population a company can select the appropriate markers that will be most beneficial for human identity testing.

Future Plans: Typing additional Y-SNPs on the U.S. population samples and evaluating the results in relation to other forensic markers: Y-STRs, autosomal STRs and mitochondrial sequence.